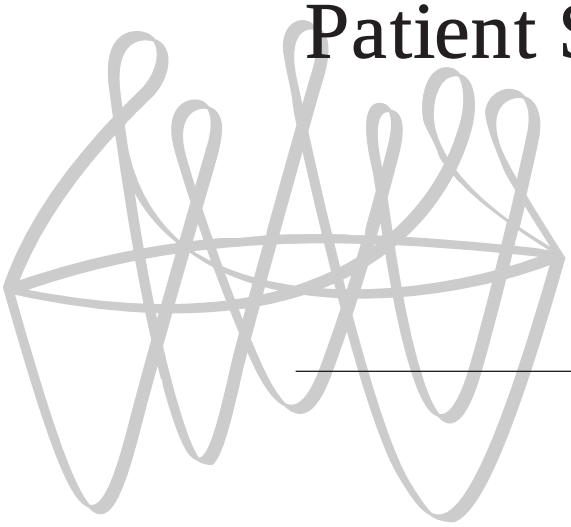
A Tale of Two Stories: Contrasting Views of Patient Safety

Report from a Workshop on Assembling the Scientific Basis for Progress on Patient Safety

National Health Care Safety Council of the
National Patient Safety Foundation at the AMA

A Tale of Two Stories: Contrasting Views of Patient Safety



Report from a Workshop on Assembling the Scientific Basis for Progress on Patient Safety

National Health Care Safety Council of the
National Patient Safety Foundation
at the AMA

Richard I. Cook
University of Chicago

David D. Woods
Ohio State University

Charlotte Miller
National Patient Safety Foundation
at the AMA

Supported by grants from:
US Department of Veterans Affairs
Agency for Health Care Policy & Research

Author information

Richard I. Cook, MD, is Assistant Professor in the Department of Anesthesia and Critical Care at the University of Chicago in Chicago, IL. He is Director of the Cognitive Technologies Laboratory. He is also a member of the Board of Directors and the Executive Committee of the National Patient Safety Foundation at the AMA. His e-mail address is: ri-cook@uchicago.edu.

David D. Woods, PhD, is Professor in the Institute for Ergonomics at The Ohio State University in Columbus, OH. He is also the Chair of the National Health Care Safety Council of the National Patient Safety Foundation at the AMA. His e-mail address is: woods.2@osu.edu.

Charlotte Miller, formerly on the staff of the National Patient Safety Foundation, was responsible for organizing the workshop and capturing the results.

The NPSF greatfully acknowledges the support of its founding and major sponsors: The American Medical Association, 3M, CNA/HealthPro and Schering-Plough Corporation.

Focus on "Patient Safety"

Patient safety is a growing concern for the public, policy-makers, and all those who are involved in the delivery of health care services. The phrase "patient safety" is, admittedly, only beginning to achieve currency within the health care community and is not yet widely used among the general public. However, concerns about patient safety have found expression in the media reports of highly publicized medical mistakes (e.g. "the wrong leg"); in medical journal articles examining error in medicine (see Lucian Leape's 1994 *JAMA* article of that title) and studies of error (e.g. work by David W. Bates et al. on medication errors); and in the organizational responses of medical, regulatory and governmental bodies (e.g. the sentinel events policy of JCAHO; the inclusion of "reducing health care errors" as a goal in the report of the President's Quality Commission).

In the midst of ever-increasing technological complexity and the massive organizational changes in health care service delivery in this country, patient safety has emerged from the ambient public discussions of quality and cost as a feature that deserves special consideration.

This new emphasis on patient safety was galvanized during the landmark conference "Examining Errors In Health Care: Developing a Prevention, Education and Research Agenda" held in October 1996 at the Annenberg Center for Health Sciences. It was an unprecedented multi-disciplinary gathering—every sector, from patients to practitioners, administrators, health plans, and regulators, plus researchers, ethicists, lawyers, risk managers and quality professionals, was represented. At Annenberg, we started to find new ways to talk and think about a subject that had never been considered so broadly and openly in health care before.

Annenberg also marked the inception of the National Patient Safety Foundation (NPSF) at the AMA, an independent not-for-profit organization devoted to assuring patient safety in the delivery of health care. NPSF is modeled after the Anesthesiology Patient Safety Foundation, which was founded in 1985.

Assembling the Scientific Basis for Patient Safety

The heightened focus on patient safety has generated a concomitant interest in learning more about the body of research on the human contributions to safety in complex systems, a body of work that has developed largely outside the health care domain. More broadly, the health care community also stands to benefit by learning how other complex, high risk enterprises—such as aviation, marine shipping, or power generation—have confronted considerable technical and political challenges in the pursuit of safe operations and public confidence.

The National Patient Safety Foundation (NPSF) at the AMA, the Department of Veterans Affairs (VA) and the Agency for Health Care Policy and Research (AHCPR) are committed to learning about, using, and adding to the established research base on safety. As a first step towards realizing that goal, the NPSF, with critical financial support from the VA and AHCPR, convened the workshop "Assembling the Scientific Basis for Progress on Patient Safety" in Chicago in December of 1997.

This report from the workshop provides a much-needed grounding in the technical knowledge relevant to patient safety—the "state of the art" of the multidisciplinary approaches that have proven productive in other domains and have only begun to be applied to research in health care. The report gives insight into what kinds of research are likely to yield interesting and productive results. Like the workshop itself, the report draws together a collection of threads to make a fabric upon which to pattern future work.

Furthermore, the workshop has helped to achieve another important goal: stimulating the participation of the safety research community in projects within health care. Individual researchers who participated in the workshop have already become involved with our organizations in a variety of ways, for example, in the design of a national patient safety system at the VA and in the launch of the NPSF's first round of research grants.

Finally, the workshop has provided the impetus for the NPSF to establish the National Health Care Safety Council. This standing body of organizational design, human factors and other experts will serve as a "technical backbone" informing all NPSF activities and as a resource for others.

Patient Safety Activities of NPSF, VA and AHCPR

Our shared commitment to cultivating a strong technical knowledge base for patient safety activities is part of a broader array of patient safety initiatives that our organizations are pursuing.

The mission of the National Patient Safety Foundation is to measurably improve patient safety in the delivery of health care. The Foundation was launched by the American Medical Association in 1997 as an independent not-for-profit research and education organization comprising a broad partnership representing consumer advocates; health care providers; health product manufacturers; employers and payers (public and private); researchers; and regulators and policy-makers. NPSF serves as the forum for a diverse group of concerned individuals to think and talk about the issues and impediments to patient safety. The NPSF seeks to be a catalyst for action and a vehicle to support change and track improvements in patient safety.

The NPSF has adopted four core strategies:
1. Promote research on human and organizational error and prevention of avoidable patient injuries in health care.
2. Promote the application of knowledge to enhance patient safety.
3. Develop information, collaborative relationships and educational approaches that advance patient safety.
4. Raise awareness and foster communications and dialogue to enhance patient safety.

The US Department of Veterans Affairs launched a public-private partnership to improve patient safety in 1997. This endeavor supports the development of a number of bold initiatives focused on implementing patient safety programs within the enormous VA health care system and making the results of those efforts available as examples that can benefit health care beyond the VA system. One example of the VA's activities is the development of a patient safety reporting system, which draws on the experience of the aviation community's successful Aviation Safety Reporting System (ASRS) and also the discussion of issues relating to incident reporting and analysis in general that took place at the workshop. The VA is also sponsoring research, pioneering system-wide implementation of patient safety interventions, and striving to create a new patient safety culture.

The Agency for Health Care Policy and Research (AHCPR), a part of the US Department of Health and Human Services, is the lead Federal agency charged with supporting research designed to improve the quality and outcomes of health care, reduce its cost, and broaden access to and use of essential services. AHCPR assists caregivers, patients, plan managers, purchasers, and policymakers by developing and disseminating practical, science-based information about the effectiveness, cost, and cost-effectiveness of health care services and alternative approaches for organizing and delivering those services. AHCPR supported Dr. Lucian Leape's pioneering work on adverse drug events that focused the nation's attention on patient safety issues. The Agency continues to support patient safety efforts in many of its research programs, including its practice and technology assessments, outcomes and effectiveness research, organization and delivery studies, and quality measurement and improvement research.

Conclusion

The NPSF, VA and AHCPR are pleased to have sponsored the workshop "Assembling the Scientific Basis for Progress on Patient Safety." We anticipate that this report from the workshop will be the first in a series that, in one way or another, have their origins in the discussion that took place during those two days last December. We hope it is useful to you and that you join us in working for patient safety.

Nancy Dickey, MD	*Kenneth W. Kizer, MD*	*John Eisenberg, MD*
National Patient Safety Foundation at the AMA	*US Department of Veterans Affairs*	*Agency for Health Care Policy & Research*

Table of Contents

Tables and Figures

Acknowledgements

A great many people contributed to make the workshop possible and successful. We would like to express our deepest appreciation and thanks to:

- Martin Hatlie, the Executive Director of the National Patient Safety Foundation (NPSF) at the AMA for his encouragement and tireless work to create the National Health Care Safety Council as a part of the Foundation and to initiate this workshop as its first activity.

- The United States Department of Veterans Affairs (VA) and the Agency for Health Care Policy & Research (AHCPR) for sponsoring and participating in the workshop, for their leadership on patient safety, and for their willingness to listen to the results from other fields in the search for progress on safety in health care.

- The many participants in the workshop who gave of their time, energy, and intellect to wrestle with the difficult questions that underlie safety in the complex and changing world of health care.

- The Board of the American Medical Association for providing the facilities to hold the workshop.

- The staff of the National Patient Safety Foundation who provided critical assistance over long hours to set up, run and document the workshop so smoothly.

Many people helped to prepare this report as a means to share the ideas and interchanges at the meeting with others. In particular, we would like to recognize the contributions of Carter Mecher, MD, Larry Goldman, MD and Jeffrey Cooper, PhD who provided valuable comments to help the shape the content of the report.

A special thanks is due to Lorri Zipperer, Information Project Manager of the NPSF who led the production and editorial process. She, along with George Kruto who indexed the material, Rosalyn Robinson of the AMA who coordinated the printing, and Karen Dangremond, of Dangremond Design, Chicago, Il. who created the layout and design of the text, handled the myriad aspects required to produce this document. Thanks Lorri for coordinating everyone and applying the right mixture of tact and forcefulness to make this report come to fruition.

Richard I. Cook *David D. Woods*

Prelude

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of disbelief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or evil, in the superlative degree of comparison only."

— Charles Dickens, A Tale of Two Cities, 1859

Exploring contrasts is a powerful means for achieving new insight. Dickens juxtaposes contrasting individual stories to tell the much larger story of the French Revolution and the Terror. The first line of the novel points out that the contrasts are striking, provocative, paradoxical, and compelling. Far from resolving the contrasts in favor of one position or the other, the novel shows how this period contained all of these qualities.

For health care at the end of the twentieth century, it is also the best of times and the worst of times, a time of paradoxes and contrasts. On the one hand, splendid new knowledge, more finely honed skills, and technical advances bring sophisticated treatments to larger and more fragile populations of people than ever before. On the other hand, media and public attention is focused on "celebrated" medical accidents—chemotherapy overdoses, wrong limb surgeries, catastrophic missed diagnoses. Stunning success and appalling failure are arrayed in contrast to each other. It is in this setting that discussions about patient safety are now taking place.

Because the sources of safety and the threats to safety remain poorly understood and because scientific research on health care safety is in its infancy, the National Patient Safety Foundation at the AMA, with sponsorship from the Department of Veterans Affairs and the Agency for Health Care Policy Research, convened a workshop in December 1997 to assemble results from the science on human performance and safety from past research in other areas.

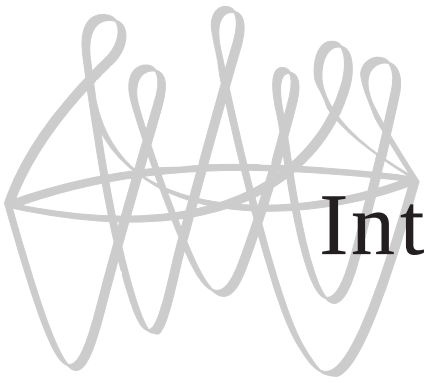
The workshop was structured around the stark contrasts between two kinds of stories we tell about accidents. Some accidents become highly visible, widely known, "celebrated" cases, e.g., the Florida 'wrong leg' case. In the first story we tell about such cases, we are puzzled. Given what we now know after-the-fact, they seem so easily preventable and the human performance so poor. We can see how the outcome could have been avoided if the people involved had just recognized the significance of some data or if they had been more careful in carrying out an activity. We fall back on explanations such as "human error" and stop, wondering how we can cope with the unreliability of the human element.

Results from close, methodical, scientific investigation of specific areas of practice in health care where failures occur (e.g., the vulnerabilities that contribute to patient injury during minimally invasive gall bladder surgery) tell a different, deeper and more complicated story. The detailed investigations are second stories revealing the multiple subtle vulnerabilities of the larger system which contribute to failures, detecting the adaptations human practitioners develop to try to cope with or guard against these vulnerabilities, and capturing the ways in which success and failure are closely related.

The second stories examine how changes in technology, procedures, and organizations, combine with economic pressures to create new vulnerabilities and forms of failure at the same time that they create new forms of economic and therapeutic success. The result is paradoxical: health care becomes simultaneously more successful and more vulnerable (or vulnerable in new ways). The changes that create opportunities and vulnerabilities also encourage human adaptation to exploit opportunity and defend against vulnerability. Individuals, teams and organizations adapt their practices and tools to guard against known threats to safety. But complexity limits the success of these adaptations. Hazards are hidden, tradeoffs difficult to assess, and the coupling across seemingly distant parts is obscured.

Digging for second stories is valuable because it promotes learning about systemic vulnerabilities. The efforts of individuals, teams and organizations to make safety are limited. People and organizations may miss or misperceive the vulnerabilities and how they come together to create paths toward failure; they may rely too much on human adaptability; they may develop brittle strategies, or they may rely on past success when change creates new challenges. How well people and organizations make safety depends on feedback to recognize systemic vulnerabilities, to evaluate the robustness of their adaptations and to understand how the changing context of medical practice affects vulnerabilities. Recognizing systemic vulnerabilities guides investments to cope with these contributors toward failure. Promoting this flow of information to learn about systemic vulnerabilities is one of the hallmarks of a safety culture.

In the workshop and in this report, the contrast between celebrated medical failures and well researched areas of human performance in medicine is used to expose the difference between the First Story of "human error" and the Second Story of systemic vulnerabilities. The different stories reveal another contrast about progress—only by constantly seeking out our vulnerabilities can we develop and test more robust practices to enhance safety.



Introduction

Increased Visibility for Patient Safety

Recent accidents in health care have fueled a growing interest in patient safety. These highly publicized accidents have occurred against a backdrop of substantial changes in the organization, delivery, and economics of health care. Together, these events lead to substantial public pressures to make progress on safety. As a result, many health care organizations are focusing more on patient safety. For example, patient safety has been the theme of major health care meetings (e.g., "Examining Error in Health Care: Developing a Prevention, Education and Research Agenda," held October 1996 at the Annenberg Center for Health Sciences, Rancho Mirage, California.), a 1998 Presidential Advisory Commission on Health Care has included patient safety as a high national priority, and major journals have recognized the topic (Leape, 1994).

The widespread interest leads immediately to two questions: "How do we make progress on patient safety in the longer term" and "What are the 'low hanging fruit' that we can pick to have an impact quickly?"

Many interested parties have widely varying ideas about the answers to these questions. Some consumer advocates desire greater public access to records about the past performance of physicians or hospitals. Other commentators promote technology as the key to progress, for example, proposing computerized physician order entry to reduce medication errors. Many want to implement more accident or 'close call' reporting systems to help identify troublespots.

How do we sort through all of these varying proposals? On what basis do we decide which proposals merit investment and change? However plausible and beneficial each proposal sounds, how can we be assured it is based on an accurate understanding of the complex factors at work? Can we distinguish approaches that will produce real progress and enhance safety from those that will lead to dead ends?

Research as a Guide for Progress

A systematic, research-based approach to the current window of opportunity on patient safety digs deeper to ask additional questions:

- What do we know about the human contribution to safety and accidents?
- How can we use this knowledge base to recognize opportunities for and obstacles to progress on patient safety?
- What does this knowledge base tell us about constructive ways to move forward?
- How do we add to this knowledge base, given unique aspects of different health care settings?

A research base on the human contribution to safety and failure in areas outside medicine has been built up over about the last 20 years. This base of knowledge has come from an intense cross-disciplinary examination of this topic driven by a series of highly visible accidents, as well as other less celebrated cases, in industries such as power generation and transportation (e.g., Three Mile Island nuclear power accident in 1979, the capsizing of the Herald of Free Enterprise in 1982, and various aircraft accidents). The participants in this work have come from a variety of disciplines including human performance, cognitive psychology, social psychology and organizational behavior, among others.

To make sense of these accidents and to develop ways to enhance safety, various researchers have collected data about the multiple human, technological, and organizational factors that contribute to accidents; investigated the normal functioning of these settings; developed new concepts and theoretical frameworks; and re-examined common assumptions. The result has been a "new look" at the human contribution to both safety and risk (e.g., Reason, 1990; 1997).

This "new look" is based on research that goes beyond the label "human error." The usual judgment after an accident is that human error was the cause, a conclusion which often serves as the stopping point for the investigation of the case. As a result, safety problems typically are seen solely or primarily as "human error" problems.

In contrast, when the label human error becomes the starting point for investigations, we find a deeper, multi-faceted story. This "second" story shows us how multiple interacting factors in complex systems can combine to produce systemic vulnerabilities to failure. The second story is more complicated but more interesting and can point the way to effective learning and system improvements.

A National Patient Safety Foundation Workshop

The National Patient Safety Foundation (NPSF) is a new organization dedicated to advancing safety in health care. Among other initiatives, the Foundation is committed to learning about, using and adding to the established research base on safety.[1] The creation of this strong technical backbone for the Foundation will not only help inform its priorities for future work, but will also help to assure that the Foundation's efforts produce progress on patient safety.

To this end in December 1997, a workshop format was used to bring together some 20 researchers and an equal number of interested leaders from health care. Each researcher is an internationally acknowledged expert in some aspect of human performance evaluation, cognitive psychology, or organizational behavior (a list of the participants is attached in Appendix A). The overall objective of the workshop, "Assembling the Scientific Basis for Progress on Patient Safety," was to develop a basis for providing sound technical advice to the National Patient Safety Foundation, a basis grounded in the research on human error, system failures, and organizational factors.

The workshop was conducted as a wide-ranging, highly informed conversation that played off the contrasts between celebrated medical failures and other cases that are less publicized but contain a significant research base on human performance. A *Sourcebook* of materials about the celebrated and uncelebrated cases helped participants prepare for the workshop (see Appendix C for a list of the articles).

The "celebrated" cases, as a group, capture the reactions of different stakeholders to medical failure. The explanations for how these cases came about are a kind of story we, as a society, tell after the fact in order to learn from the failure and to decide what kinds of changes are needed. In telling that story stakeholders focus on a few of the factors and actors that could be seen as contributing to the sequence of events. Which factors and actors come to be regarded as most responsible depends on the kind of stakeholder, common beliefs about the role human performance, common beliefs about why systems succeed and fail, and the normal human processes for attributing causes to surprising events. The story that results represents a model of the threats to patient safety and presumes that certain changes will address or eliminate these threats. This treatment represents the "first" story.

In contrast, another set of cases, uncelebrated but well researched, reveal a "second" story. This story captures how the system usually works to manage risks but sometimes fails. When researchers pursue the second

[1] The Foundation has established the National Health Care Safety Council, comprised of safety experts drawn from various disciplines and domains to help accomplish this goal. This workshop provided the first opportunity for assembling the kind of expertise that will comprise the safety council.

story they broaden the scope of inquiry in ways that lead them to identify systemic vulnerabilities that contribute to failures. The result is a very different view of the patient safety landscape, a view that highlights many factors that the first story ignores.

Each uncelebrated case yields interesting results but together they have broad implications for how to make progress on patient safety. Analyzing the uncelebrated cases as models of patient safety research represents a substantial departure from the usual, first story based approaches. The concepts and methods used in the research for these cases can serve as a model for approaches to other areas in health care.

The workshop also examined lessons from incident reporting and analysis activities in other domains. This topic often is cited as a key initial step toward enhanced safety in health care. The departure point for this portion of the workshop was a presentation on the lessons learned as aviation safety experts sought to develop a system for collecting and analyzing incidents. An edited version of this brief, but powerful presentation is attached in Appendix B.

The workshop used the contrasts between first and second stories to evoke participants' comments about research on patient safety. The result was a dynamic and complex exchange. The discussion did not point to any simple answers. Rather it produced some strong indications of where and how fundamental progress can be made. It also showed how some of the approaches health care organizations often adopt have proved to be of limited value in other settings.

The contrasts between these two kinds of stories about patient safety also provide the structure for this report. The report conveys much of the flavor of being there, but does not attempt to reproduce the entire discussion. Instead, it represents the authors' construction of the central themes and concepts that emerged from the discussion, based on an analysis of the verbatim transcript of the meeting and the sourcebook materials. Far from being the final word on these important topics, we hope the contrasts captured in this report will broaden other discussions of safety and ground these debates in the basic results from past research.

Health Care After Its "Three Mile Island"

At the beginning of the workshop, an analogy was made between the state of health care today and conditions in the nuclear power industry in 1979 after it was staggered by the Three Mile Island accident.

The Three Mile Island accident was a watershed for the nuclear power industry. It irrevocably changed the way people looked at nuclear safety, and this change created both the possibility of and the need for new approaches to safety in that industry.

While there is no single medical event comparable to the Three Mile Island accident, the combined effect of the celebrated medical failures over the last few years is similar. The attention these cases have received, and the debate and action they have engendered, signal a fundamental shift in public perceptions of patient safety. Although the cases are spread out geographically and involve different kinds of failures, in combination they have shifted the public perception of the sources of risk and failure in medicine. The founding of the National Patient Safety Foundation (NPSF) itself is a marker for this change, as are recent initiatives from regulatory, advisory, and legislative bodies.

Health care stands in 1998 where nuclear power stood at the end of 1979. There are growing public demands to enhance patient safety. There is public concern that the financial pressures and organizational change in health care will degrade practitioners' expertise, create conflicting goals and incentives, increase workload, and reduce safety margins. There are concerns about the nature of training and certification of practitioners and institutions. There are anxieties about injuries from a panoply of technological devices, drugs, and techniques. Not everyone is concerned, nor are all in agreement about the sources of hazard or the appropriate responses. There is confusion and argument about the meanings of events. In this, there is a close parallel to the time just after the Three Mile Island accident.

This situation is fraught with promise and also with risk, hence, "it is the best of times, and the worst of times" to explore safety in health care.

Despite the fact that celebrated and uncelebrated cases underscore medicine's fallibility, health care today is more technically advanced than it has been at any other moment in history. But the ever advancing state-of-the-art of medicine has been coupled with ever increasing complexity. Against a backdrop of organizational change and economic pressures, the increasing complexity of health care increases the possibilities for unanticipated and unintended consequences. As a result, even more opportunities for failure may exist.

The intense interest generated by medical accidents may provide opportunities to advance patient safety. Political will and economic investment follow public attention and can provide the energy to implement meaningful change in health care. On the other hand, there is a downside to this sort of public attention. The need to do something, to react quickly, to provide visible (if not substantive) evidence of progress, may result in a rush to implement unproductive or counterproductive programs. It may even result in direct efforts to manipulate the image of safety to promote political or economic interests.

Defining Opportunities and Obstacles

In this context, we asked the gathered experts on safety related issues to help us take the existing research base as a guide for how we could move forward on safety in health care.

To this end we posed a series of safety-related questions to the participants:

- What lessons can we learn from studies of failure and success in other domains?
- What scientific knowledge is available regarding the human contribution to risk and safety?
- Given the scope and complexity of health care, the diverse collection of issues that influence individual health care practitioners and coordination across health care teams, technological factors, organizational context, and regulatory pressures—What are the opportunities and obstacles for making progress on patient safety?
- What can the research base teach different stakeholders in health care about the factors that produce failure?
- What can the research base tell us about the kinds of investments and changes that will enhance safety?
- What cautions or warnings about unproductive or counterproductive approaches are needed?
- What meaningful guidance can we provide about the priorities for future research and applications?

Day One – Contrasting Cases

For the first day of the workshop, the discussion was organized around specific cases of celebrated accidents and uncelebrated areas of research related to patient safety. These cases served as a framework within which to elaborate issues, opportunities, obstacles and perspectives on failure.

Celebrated Accidents

The "celebrated" cases are medical accidents that have attracted a great deal of attention from the public and the press (Table 1; Figure 1). The reports of these cases have led to a variety of reactions from health care professionals, regulators, and the public. Interestingly, many of these cases have achieved such a level of prominence in the collective public psyche that one can elicit a collection of images simply by mentioning the "Florida wrong leg or Willie King case" or the "Betsy Lehman case" or the "Libby Zion case." All of these cases evoke our empathy for some tragic loss.

These cases also have become symbolic in other ways. The case of Willie King in Florida, in becoming the "wrong leg case," captures our collective dread of wrong site surgery. The death of Libby Zion has come to represent not just the danger of drug-drug interaction but also the issues of work hours and supervision of residents – capturing symbolically our fear of medical care at the hands of overworked, tired, or novice practitioners without adequate supervision. Celebrated cases such as these serve as markers in the discussion of the health care system and patient safety. As such, the reactions to these tragic losses become the the starting point for discussions of obstacles and opportunities to enhance safety.

Table 1: "Celebrated" medical accidents.

Celebrated Case…	…an example of…
Florida wrong leg case (the Wille King case)	Wrong limb/site surgery (symmetry failure)
Betsy Lehman case	Chemotherapy overdose
Gargano case	Chemotherapy overdose
Ben Kolb case	Epinephrine/local anesthetic solution swap
Libby Zion case	Drug-drug interaction
New York doctor (Einaugler case)	Injection port identification failure (criminalized)
Colorado nurses case	Route of administration failure (criminalized)

Note: Since the conference, there have been reports that the surgeon involved in the "wrong leg" case has been involved in a "wrong patient" case. Again, there is a first story that focuses on the individual, but also glimmers of a deeper second story that could reveal more about general vulnerabilities to this kind of failure.

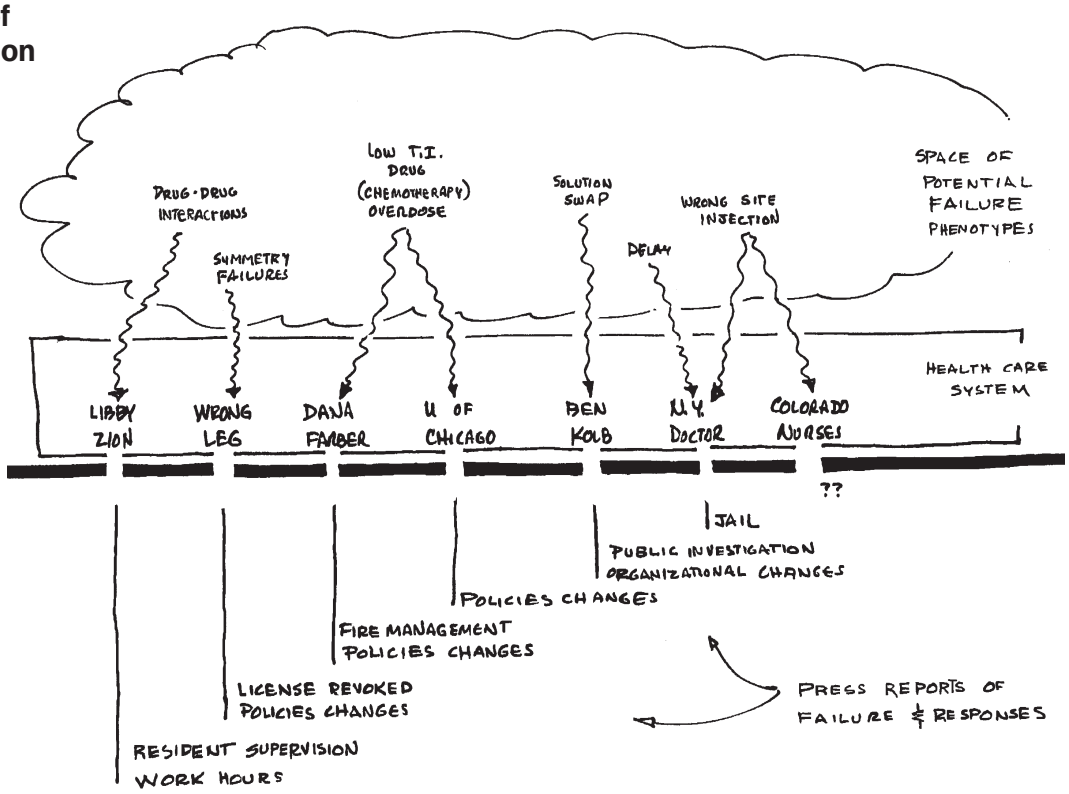
Since the conference, the criminal proceedings against the Colorado nurses have been resolved (see *ISMP Medication Safety Alert!* Volume 3, Issue 3, February 11, 1998 from the Institute for Safe Medication Practices at ismpinfo@ismp.org).

The sourcebook distributed to participants contained background on a selection of these cases. Most of the available material comes from newspaper articles on a specific case. Also included were two broader perspectives from reporters looking across multiple health care accidents and attempting to synthesize a more coherent picture of accidents in general (Lisa Belkin, *New York Times*, June 15, 1997; Steve Twedt, *Pittsburgh Post-Gazette*, October 24-31, 1993).

The sourcebook also included a column by the news commentator Sidney Zion, father of Libby Zion. This commentary appeared following a media briefing conducted by the National Patient Safety Foundation. It conveys some of the charged atmosphere surrounding public discourse on patient safety issues. It may presage the sorts of communication difficulties that will confront those who seek to develop and explore a more technically grounded view.

Typically there are no independent investigations of the sequence of events and contributors to the outcome such as those done by the National Transportation Safety Board (NTSB) following aviation accidents. As a result, we often must rely on news reports, but other sources of information may be available. In the "wrong leg" case there is also an official document prepared by a Florida hearing officer. This provides the legal rationale for the revocation of the surgeon's license and gives an account of the chain of events leading to the amputation of the wrong leg. The materials for the Ben Kolb case include a set of statements giving the perspectives of the hospital CEO, the risk manager, the physician, the medical liability insurer, and the family.

Figure 1. The view of patient safety based on celebrated cases.



The View of Patient Safety from Celebrated Cases

The First Story

Each of the celebrated cases consists of a relatively straightforward, simple, and easily understood story constructed after the accident. This first story typically explains the accident in terms of a simple cause, usually human operator error. The people held responsible are those closest in time and space to the final outcome, especially those who, it is believed, could have acted in another way that would have led to a different outcome. Their behavior, with knowledge in hindsight of the tragic outcome, appears to be outrageous, willful disregard of critical cues or factors. In retrospect, each failure seems preventable by relatively simple means, such as new policies and procedures or calls to increase the "vigilance" of practitioners. Finding the culprit ends the investigation.

Each celebrated case has come to represent a specific threat to safety for the public and for people in health care. Figure 1 shows schematically the relationship between the celebrated cases and the sources of failure they seem to represent. However, the reactions in these celebrated cases provides only a partial view of the hazards and of the factors that contribute to failures.

In the "wrong leg case," for example, public attention focused on the surgeon. However, this case represents a larger class of failures that arise because of the symmetry of the human body. Bilateral symmetry (paired organs, limbs, etc.) creates the inherent risk of wrong site, wrong side, wrong limb failures. This risk is well known. Health care practitioners and organizations recognize this risk and have a variety of defenses against just this sort of accident. Cases like the Florida wrong leg amputation are situations where all the defenses broke down or were ineffective. They point, not so much to inadequate defenses, as to a systemic inability to maintain these defenses in working order in the face of a variety of pressures. Indeed, the accident is a potential source of data about the larger system that delivers care, not simply about a flawed individual. Understanding the factors that bypassed or undermined the defenses in this case would help us learn in ways that could be applied to other situations, practitioners, and organizations.

The response to failure is the most significant feature of each of the celebrated cases. The affected organization usually creates new policies or procedures in the hope of forestalling any repeat of this particular accident. For example, several celebrated cases have involved medication misadministrations. Responses have included efforts to more tightly control the use of the particular drugs by imposing the requirements for more elaborate checks and additional steps in prescribing and dispensing medicines.

However, some drug misadministrations point to a broader risk related to drugs with a low therapeutic index. These are drugs where the effective dose is near the toxic dose. Recognizing this class of drugs and associated risks can help illuminate weaknesses in the defenses deployed to mitigate those risks. Drug-drug interactions, e.g. the Libby Zion case, represent a hidden low-therapeutic index situation that contributes to the outcome in complex ways.

The response to the failure also provides significant data about how attributions of causality are made and how responses to a widely publicized failure are driven by a variety of factors. The limited ability of the regulatory bodies to influence safety in a direct way is one remarkable characteristic of the 'wrong leg' accident. The regulators' choice to "send a message" by revoking the surgeon's medical license is itself an interesting feature of the case. The decision to send such a message and phrase it in this fashion signals a set of beliefs about which factors lead to failure, which interventions can change those factors, and how health care practitioners are expected to react to these messages. These beliefs constitute one model of how people contribute to safety and risk. The scientific question is, do these beliefs correspond to accurate models of the factors that affect success and failure and of the factors that enhance or degrade the performance of health care practitioners?

Taken together, the celebrated cases were tantalizing to the researchers. They indicate the potential for catastrophic failure in health care. They also demonstrate the way that these failures can capture public attention, evoke outrage, and provide impetus for regulatory action.

But the cases are also remarkable for their limitations. The workshop participants rapidly tried to move beyond this "first" story into a discussion of the deeper "second" story that lies behind such cases. However, in each instance, the story told after the event is too simple. Its details are too limited to serve as the basis for understanding the interplay of the multiple contributors that led to the accident. While the accounts of the celebrated failures do tell us a great deal about the social response to failure, the participants observed that the celebrated cases do little to broaden our understanding of the other risks that exist in health care, the sources of these vulnerabilities, or the means for reducing them.

To better understand systemic vulnerabilities to failure and to see how failure is usually prevented requires collecting a different kind of data, analyzed in different ways, that reveal a different story. To accomplish this requires analysis based on concepts grounded in the research base about the factors that affect the many different kinds of human performance relevant to health care settings. In particular, the researchers recognized the impact of hindsight bias on the construction of these first stories of accidents.

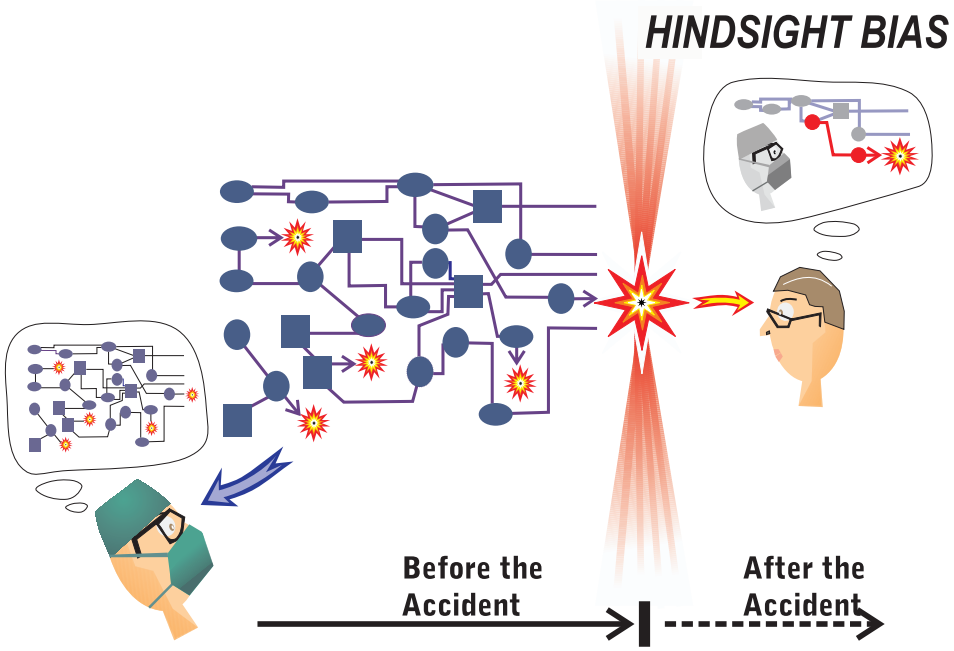
The Hindsight Bias

The tendency to attribute accidents in health care to simple causes such as isolated human failures is derived in part from a particular form of bias that clouds post-accident reviews of human performance. It is well documented that knowledge of outcome biases our later judgments about the processes that led up to that outcome (See Figure 2). The way we look back is shaped by the outcome. That outcome knowledge, however, was not available to the participants before the fact. In looking back we tend to oversimplify the situation the actual practitioners faced. This blocks our ability to see the more complicated, richer story behind the label human error.

The hindsight bias effect is a well reproduced research finding relevant to accident analysis and reactions to failure.[2] In the typical study, two groups of judges are told a story and asked to evaluate the performance of characters in the story. The story is identically told for each group of judges with a single exception: the difference is the *outcome* of the story. One group is told the episode ended in a poor outcome (death, significant

[2] See Woods et al. (1994), chapter 6 for an overview of the research. For the original studies see Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, I,* 288-299 and Baron, J. and Hershey, J. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*, 569-579. For a replication in medicine, see Caplan, R., Posner, K., and Cheney, F. (1991). Effect of outcome on physician judgements of appropriateness of care. *Journal of the American Medical Association, 165*, 1957-1960.

Figure 2. Hindsight does not equal foresight. Knowledge of outcome biases our judgment about the processes that led up to that outcome.



loss, etc.). The other group is told that the outcome was good (minor injury, insignificant loss or even some gain). The two groups of judges consistently differ in their assessment of the story characters' performance. Judges told of the bad outcome assess the performance as flawed. Judges told that the outcome was successful assess the performance as acceptable. The differences are stark, repeatable, and strong. In fact, hindsight bias impacts judgments even when judges are warned that the outcome knowledge may influence their ability to make assessments.

It is clear that hindsight bias poses a great obstacle to understanding patient safety through celebrated cases. The powerful outcomes of these accidents shape the way that post-accident (looking back) judgments of human performance are made. This bias limits the value of these cases because the debris of outcome obscures the complexity of the situation confronting practitioners. This leads to simple "first stories" of accidents and, paradoxically, limits what can be learned about safety from such events.

Uncelebrated Cases: The Second Story

The next phase of the workshop provided a contrast to the celebrated cases in the form of "uncelebrated" cases where some research base was available. The cases highlight some of the factors that affect success and failure in the practice of medicine. More importantly, the cases demonstrate the kinds of factors that affect the success and failure of research on human expertise and its role in system performance. In contrast to the celebrated cases, a multi-faceted story about how the system works and how it sometimes fails unfolds in the investigations.

In this section, we provide a summary of what was presented about the three uncelebrated cases. Then we draw out some of the larger implications of each case for patient safety in general based on the discussion at the workshop and the research results themselves.

While the details of the uncelebrated cases are distinct, they have at least five features in common.

First, each uncelebrated case shows that bad outcomes flow not from single-point failures but from a set of factors. The research reveals that these factors are each necessary but only jointly sufficient to cause an accident. The analysis in the uncelebrated cases exposes the system issues and latent factors that contribute to failure.

Second, these investigations show that enhancing safety begins with efforts to understand not just the sources of failure but also the sources of success. System operations are seldom trouble-free. In every close examination of complex systems in operation, observers find many more opportunities for failure than actual accidents. The difference between the high potential for failure and the low rate of failure is produced largely by practitioners.

Much of expertise and skill is directed towards preventing poor outcomes or recovering from problems before their consequences impact on the patient. Each investigation shows how practitioners resolve conflicts, anticipate hazards, accommodate variation and change, cope with surprise, work around obstacles, close gaps between plans and real situations, detect and recover from miscommunications and misassessments. In these activities practitioners regularly forestall or deflect potential accident trajectories.

Put another way, human practitioners are not so much the cause of occasional sporadic accidents as they are the active agents that regularly contribute to success. When they carry out their roles successfully, they are the active creators of safety. Safety research tries to identify factors that undermine practitioners' ability to do this successfully.

Third, the research results shift attention away from the people closest to the accident and toward the blunt end of the system where regulatory, administrative, and organizational factors reside. Complex systems such as health care or aviation have both a sharp end and a blunt end (Figure 3). The sharp end is where practitioners interact directly with the hazardous process in their roles as pilots, mechanics, air traffic controllers, and, in medicine, as nurses, physicians, technicians, pharmacists and others. At the blunt end of the health care system are regulators, administrators, economic policy makers, and technology suppliers. The blunt end of the system is the source of the resources and constraints that form the environment in which practitioners work. The blunt end is also the source of demands for production that sharp end practitioners must meet. The

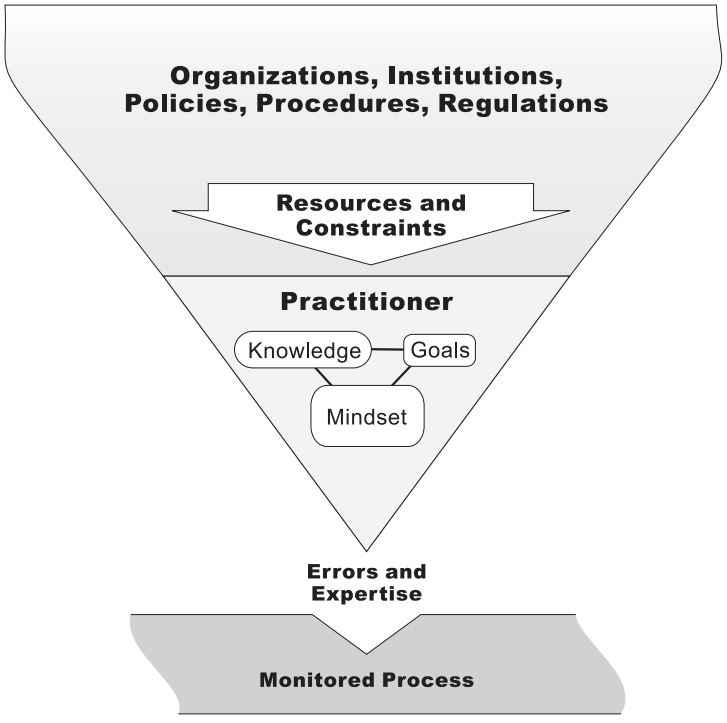
demands are often conflicted, as when the blunt end provides incentives for greater production while simultaneously demanding lower rates of failure.

The more safety researchers have looked at the sharp end, the more they have realized that the real story behind accidents depends on the way that resources, constraints, incentives, and demands produced by the blunt end shape the environment and influence the behavior of the people at the sharp end (Reason, 1997). Detailed examination of accidents in these systems consistently shows that the ability of sharp end practitioners to defend against failure in these cases depended directly and indirectly on a host of blunt end factors rather than on the isolated "error" of human practitioners.

Fourth, the research methods described in reviewing the uncelebrated cases demonstrate, in part, of how research that ultimately improves safety is done. The combination of methods chronicled in these cases may be unfamiliar to many, but they represent the kinds of techniques that have been developed to understand the role of human performance, human-machine cooperation, and cooperative work in complex evolving situations.

Fifth, and perhaps most important, the research on the uncelebrated cases points to areas where substantial progress can be made. Significantly, these are not single, local fixes or "magic bullets." Rather, the research reveals a set of factors involved in failure and shows that there are multiple directions for improvements that need to be coordinated in order to make progress on safety.

Figure 3. The blunt end of a complex system controls the resources and constraints that confront the practitioner at the sharp end.



Modified from Woods, et al. 1994

Uncelebrated Case #1: Bile duct injuries during laparoscopic cholecystectomy[3]

The shift to laparoscopic cholecystectomy where surgeons use a small video camera to indirectly observe the process of removing the gallbladder, has been accompanied by an increase in bile duct injuries with significant consequences for patients. Studies of the basis of surgical expertise at this task revealed the need for new critical perceptual and cognitive skills, in particular, a kind of judgment under uncertainty and risk when considering to convert to an open procedure if the anatomy cannot be clearly visualized. The research results identify opportunities to improve performance through new perceptual aids, new techniques for training judgment under uncertainty, and needed changes in organizational behavior.

Background

Technological change sometimes brings new problems that demand attention. The growth of laparoscopic cholecystectomy provides an excellent example. This form of minimally invasive surgery to remove the gallbladder has largely replaced the older, "open" cholecystectomy. The laparoscopic procedure involves use of a small video camera to provide a view of the gallbladder and surrounding structures as these are manipulated with instruments that penetrate the abdominal wall. The procedure involves a few small incisions rather than one large one and, for this reason, allows faster recovery, shorter hospital stays, and less pain than the older "open" technique.

The widespread adoption of laparoscopic cholecystectomy provided obvious benefits to many, but it was accompanied by a significant increase in incidence and severity of injury to the common bile duct (Way, 1992). Injury to this structure, which carries bile from the liver to the intestine, is a catastrophic accident that can lead to protracted hospitalization, multiple surgeries, and even liver transplantation. Bile duct injuries are not unique to laparoscopic cholecystectomy; they can occur even with the open procedure. But decades of experience with open cholecystectomy had reduced the rate of bile duct injury to a very low level.

The new rash of bile duct injuries associated with laparoscopic cholecystectomy was troubling, especially because the benefits of laparoscopic surgery were so compelling and the demand for this form of cholecystectomy was intense (Way, 1992). It was also troubling because the severity of the injury increased. Those suffering bile duct injury during laparoscopic cholecystectomy tended to have devastating injuries of the sort that might lead eventually to liver transplantation. Furthermore, bile duct injury was frequently compounded when the original surgeon attempted a repair procedure. The repair was more likely to be successful if it was carried out by a specialist with experience at biliary duct repair (Stewart and Way, 1995).

[3] See Dominguez, Flach, Lake, McKellar & Dunn (in press); Dominguez (1998); Way (1992); Stewart and Way (1995).

The increase in severity and frequency of injury was detected fairly early during the expansion of laparoscopic cholecystectomy. Epidemiological work suggested that surgeon inexperience played an important role; surgeons less experienced with this technology had a higher rate of injury. Clearly the remote viewing and manipulation of the tissues that goes along with the laparoscopic technique altered the mix of optimal surgical skill in ways that increased the risks for injury.

Experts with laparoscopic technique have pointed out that minimizing the risk of bile duct injuries depends on identifying the bile duct anatomy before cutting, clipping, or cauterizing any structure. However, bile duct injuries occur only in the setting where the operating surgeon thinks that the anatomy has been identified — no one cuts the common duct knowing that it is the common duct!

Ideally, practitioners should take steps to definitively identify the anatomy when the structures (ducts and arteries) can not be identified clearly. One option is to convert the procedure from laparoscopic to open to permit direct handling of the tissues and a direct binocular view of the anatomy, but this decision sacrifices the advantages of the laparoscopic procedure. The decision to convert is a new judgment under uncertainty and risk. The research has explored the factors that affect the difficulty of making this judgment and what constitutes expertise at this judgment.

The Research

Dominguez and colleagues investigated the nature of surgical expertise during laparoscopic cholecystectomy, in particular the judgment to convert to an open procedure.

To study this judgment they had to create conditions where visualizing the anatomy is challenging. A challenging case is needed to observe how surgeons evaluate whether to continue laparoscopically or to convert to an open procedure. Surgeons do not confront this decision at a single, well-defined moment. Rather the issue emerges over time as the physician explores the situation and confronts difficulties. Thus, Dominguez et al., used a complicated case that included a number of difficulties that interacted and fed upon each other.

Because laparoscopic surgery is done using video displays of the surgical field, they used a videotape of a difficult procedure to present the case to 20 surgeons and surgical residents. For this type of surgery, the videotape record contains precisely the same visual information that was available to the surgeon who actually did the operation, presented in precisely the same way.

As each surgeon viewed the video of the surgical field as the case progressed, they commented on the nature of the case, the probable future

course and, specifically, the comfort level they were experiencing with continuing the case laparoscopically, as opposed to converting to an open procedure. The investigators also stopped the video at specified points to ask the surgeon questions on his or her assessment of the situation.

The study, like any other exploration of people encountering and coping with real problems, produced a data set that was challenging to analyze. The basic results are the step by step assessments of each participating surgeon as the procedure evolved. These are built up from the comments of the participants as they viewed the videotape, linked to the characteristics and difficulties of the case as it unfolds moment by moment. This kind of data analysis, called protocol analysis, examines the process by which someone solves a problem. It is and has been the basic technique used to study problem solving. Dominguez et al.'s protocol analysis was made richer by including data from a variety of practitioners with varying degrees of experience.

A variety of results emerged from analysis of the surgeons' commentaries on the videotaped cases. Basically, it reveals that the conversion decision is a difficult tradeoff judgment. The data show some of the visual cues that trigger consideration of whether to convert to an open procedure. The results provide insights about how laparoscopic visualization is limited when compared to direct binocular vision and handling of tissues. The results also provide some insight into the circumstances in which people may make the tradeoff inappropriately, for example, by continuing the procedure laparoscopically even in the face of increasing uncertainty.

Significantly, the research goes beyond merely identifing visualization as a key factor. It does much to define *what* needs to be enhanced, *where* the critical visual cues reside, and *how* the cooperative work of the surgical team is organized around visualization. These kinds of results are important because they suggest different interventions that can improve performance. Information about difficulties in visualizing the anatomy, combined with knowledge of human perception, suggests that perceptual aids to enhance surgeons' ability to visualize the anatomy through a two-dimensional video view of the surgical field would be valuable.

The method used by Dominguez et al. also provides insight into how training for surgeons should be modified to prepare them for the new judgments that new technology demands. By confronting videos of cases specifically chosen to display the variety of factors that play into the decision to convert to an open procedure, surgeons can expand their expertise. This technique, sometimes called exploratory learning, is now being used with simulation technology to train high-performance skills (e.g., Feltovich, Spiro and Coulson, 1989; Howard, Gaba, Fish, Yang, and Sarnquist, 1992).

While Dominguez's study is closely focused on perceptual and cognitive factors, the results have much broader implications. When patients, referring physicians, administrators, and colleagues discover that a case has been converted from laparoscopic to open, their response influences future surgical decisions. This is a specific example of the way reactions of larger organizational and professional groups to practitioner decisions in specific cases has a strong influence on the way people make judgements in the face of uncertainty and risk. Organizational responses to cases are an important part of the practitioners' world and encourage them to adjust their approach to risk.

Indeed, this organizational response becomes one point of certainty in a world where future outcomes are inherently uncertain. The core issues of bile duct injury in laparoscopic surgery are mainly how individuals and groups cope with uncertainty. While they did not study these factors directly, Dominguez et al.'s results show that it is impossible to formulate narrow, rule-based approaches to the problem of bile duct injury in laparoscopic surgery. They show that injuries arise from the same sources that produce the (usual) success of this method.

Finally, the technological change represented by laparoscopic techniques raises new questions about skill that have organizational and professional implications. As a new generation of surgeons emerges, will they have experience only with laparoscopic techniques? Will they be reluctant to convert even in cases where uncertainty is high? The skill mix changes as technology changes. This has profound implications for training, especially for more difficult or complex situations. These are the same issues raised by Way in his editorial. They are especially important in an environment where there are substantial pressures to reduce training time and costs, to reduce the skills required of practitioners, and to increase production.

Dominguez et al. provide a model for exploration of these issues that can be extended and reused. Their research is not simply a study but also the model for a host of studies that can explore the complexity and uncertainty of surgical decision making, expertise, and injury.

Implications of the Research

This uncelebrated but researched case is interesting not simply as a specific area in the landscape of patient safety, but also as a model that illuminates broad generic issues.

This is an excellent example of the way periods of significant or rapid technological change create demands for new skills and judgments. These demands can contribute to new kinds of failures with new consequences for failure. In the case of laparoscopic cholecystectomy, there is a path toward failure that did not exist before the new technology (i.e., failing to

convert to an open procedure when uncertainty is high). The consequences of failure are changed, too: bile duct injuries, when they occur, are more likely to be severe. This same pattern of new technology, leading to new demands, leading to new forms of failure with altered consequences, has also been predicted for infusion-based total intravenous anesthesia (Cook and Woods, 1996).

The case of bile duct injury during laparoscopic cholecystectomy also shows the importance of pursuing the second story that lies behind the first, superficial story of isolated practitioner failure as the source of accidents. This deeper look identifies a set of factors that combine to produce *both* success *and* failure. Perceptual, cognitive, and organizational factors all play roles in this case. Reducing the rate of failure involves improving the system which depends on a coordinated approach that develops and evaluates (1) perceptual aids, (2) exploratory learning techniques to enhance high performance skills and expertise, and (3) changes to organizational behavior. This requires investments of time, energy and financial resources.

Grappling with an area like laparoscopic cholecystectomy means developing an understanding of how practitioners handle uncertainty, risk, and hazard. In this, researchers are confounded by hindsight bias. Uncertainty exists only so long as the outcome is undetermined. Hindsight bias tends to make it hard for us to appreciate the uncertainty practitioners confront. It is easy, when there has been no bile duct injury, to see the decision to convert to open cholecystectomy as too conservative, as sacrificing important goals when no such sacrifice was necessary. In hindsight, we readily identify practitioners as risk-averse or risk-seeking. But in actuality, risk is an inherent part of their world, a fluid and changing characteristic that can be difficult to localize and is impossible to quantify. Given the very high consequences of bile duct injury, handling the conversion tradeoff well means that sometimes surgeons will convert even though hindsight will reveal it was probably unnecessary. Some patients will be harmed (suffer the undesirable effects of an open cholecystectomy) so that others may benefit (avoid bile duct injury). Although Dominguez et al. began with a relatively narrow study of perception and visualization in laparoscopic cholecystectomy, their results provoke consideration of much larger issues, making a rich network of connections with other research results.

The Dominguez et al. study illustrates how one can proceed to enhance safety in other areas of health care. First, they looked at the sources of both success and failure. They began by studying what makes problems more or less difficult. This helped them identify the human performance issues relevant to expertise (e.g., perceptual factors and judgment under uncertainty and risk). Significantly, they recognized that practitioner performance depended on the larger organizational context.

Getting the results depended on tracing the process of how practitioners handle different kinds of situations. Getting this story, in the form of a problem-solving protocol, is necessary in order to learn about human contributions to risk and safety. Researchers then can look for and tabulate patterns across these problem-solving protocols.[4] When these methods are used, investigators begin to escape from hindsight bias to find the set of multiple interacting factors that contribute to accidents.

Uncelebrated Case #2: Antibody misidentification and transfusion reactions[5]

Smith and colleagues studied the factors that contribute to antibody misidentification. The studies of expertise led to the development and testing of new systems to aid lab technician performance on this task and to support improved training. Tests of the new system have shown substantial decreases in antibody misidentification.

Background

There are a number of antigens that may be present on red blood cells. These can cause a transfusion reaction if a patient sensitive to these antigens receives blood that contains them. To avoid these reactions, lab technicians check to see which antibodies are present in a patient's blood prior to transfusion.

Blood screening checks for the presence of a variety of antibodies in the patient's blood using the results of a battery of tests performed in parallel. Taken together, the results of these tests indicate the types of antibodies present. The results are presented in a table that relates reactions to hypotheses about what antibodies may be present. The technicians evaluate the pattern of reactions and make inferences about which antibodies are present. Knowing which antibodies are present allows them to select blood units without corresponding antigens for transfusion. A crossmatch is then performed using the selected units of blood and the patient's own serum in order to verify the compatibility.

The research project was undertaken to develop improved means to train technicians for this task. In particular, the goal was to use new technology (e.g., artificial intelligence) and techniques (e.g., exploratory learning) to develop computerized tutor and learning aids.

The Research

The project was not driven by reactions to visible or celebrated failures, i.e., transfusion reactions leading to severe patient consequences. Instead, the motivation was a desire to demonstrate the use of new information

[4] In problem-solving research, the term "protocol" traditionally refers to a description of the process by which a problem is detected, framed, investigated and resolved. Medicine uses "protocol" to refer to a procedure or guide for treatment.

[5] See Obradovich, Smith, Guerlain, Rudmann, Strohm, Smith, Svirbely, & Sachs (1996) and Guerlain, Smith, Obradovich, Rudmann, Strohm, Smith, & Svirbely (1996).

technology and to assist the training of new technicians. The research team began with studies designed to examine the ways that experts perform the task and to explore the contrast between strategies used by experts and the behaviors of students and less experienced practitioners.

Multiple research methods were used, all focusing on (1) identifying what makes problems difficult and then (2) using difficult problems to understand what characterizes successful and poor problem-solving strategies. The methods used included critical incident analysis, knowledge elicitation based on walkthroughs of cases, observation of practitioners solving real problems in actual facilities, and observation of practitioners solving simulated problems using a high fidelity computerized test bed for exploring new approaches for support and training.

The research showed that failure occurs even with easy tasks because the tools people use create vulnerabilities. Antibody misidentification can arise from "slips" traced to characteristics of the paper tools used for record keeping. Given the identical rows and columns of the current paper format (Figure 4), it is relatively easy to start reading across a row and shift up or down so that the wrong row is scanned. This leads to a misinterpretation of the pattern of reactions. This link between the design of the paper forms and a form of failure suggests some straightforward interventions to improve performance. Thus the research points to some of the "low hanging fruit" that safety advocates believe should be exploited quickly to improve safety. It is worth noting, however, that being able to see this sort of opportunity is easy in hindsight, after the research is done—it was not recognized as sigificant before.

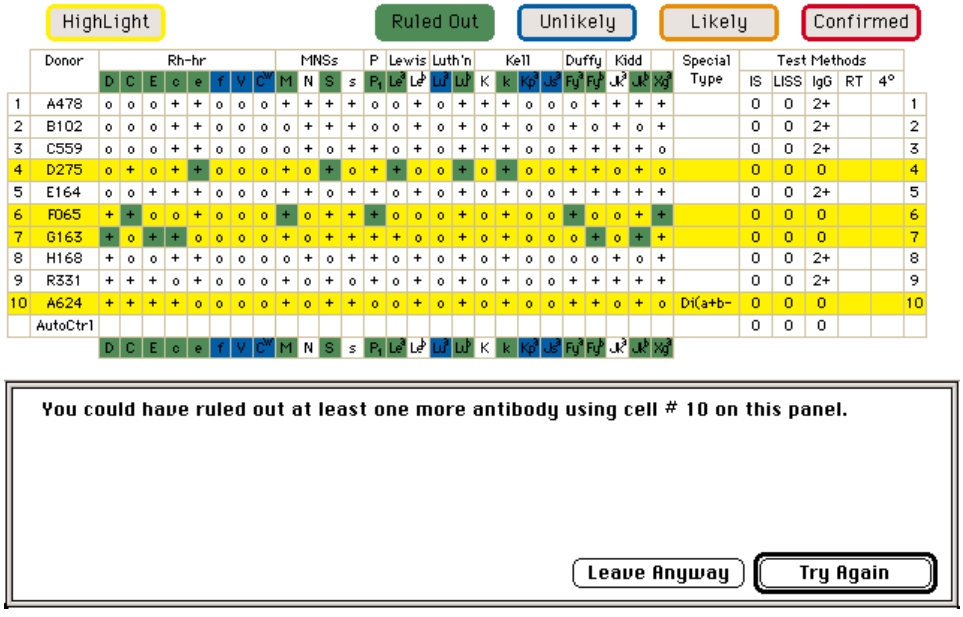


Figure 4. A stage in an antibody identification problem using an enhanced electronic version of the original paper form with computer-based critiquing.

Some cases of antibody indentification are difficult because they contain factors such as noisy data, weakly reacting antibodies, or multiple interacting antibodies that mask each other. Cases with these attributes often proved very difficult for both students and practicing lab technicians, and they frequently went "down the garden path."[6] Misidentifications were surprisingly frequent for some cases with these characteristics with rates approaching 50% among practicing lab technicians.

Other problems are difficult to solve because they challenge the fund of knowledge that lab technicians possess. As in other areas studied (Feltovich, Ford and Hoffman, 1997), practitioners sometimes possess particular misconceptions that lead to poor performance on certain kinds of problems. For example, technicians sometimes had a misconception regarding the effects of pre-warming on reactions.

As these studies examined expert strategies, it quickly became apparent that more experienced practitioners had developed strategies that were sensitive to possibility of misidentification. Some were generic strategies that helped avoid traps or recover from a tentative misidentification. Others were tailored to help avoid specific vulnerabilities. Less expert practitioners did not possess these strategies for detecting and recovering from incipient misidentification.

Finding these kinds of results depended in part on knowing where to look. Past research on human performance on diagnostic tasks suggested that there would be classes of problems that offer subtle or infrequently encountered patterns that are more likely to lead to misidentifications. The investigators, as they learned more about which basic patterns were embodied in the antibody identification task, were able to predict where misidentifications would occur.

As the researchers began to understand what made certain kinds of problems difficult and how expert strategies were tailored to these demands, the researchers asked certain questions. What kinds of tools could be used to assist lab technicians? What kinds of training could improve their knowledge and strategies for difficult cases? What would help a lab technician recognize that a particular case was likely to be difficult? How could lab technicians decide when they needed help from more experience personnel to solve the case they faced?

Following these explorations, the simulation of antibody identification tasks became a test bed for exploring the impact of different strategies to

[6] A generic class of problems have been termed "garden path" in research on problem solving because the pattern of initial evidence makes it easy for people to focus and become fixated on a plausible but erroneous diagnosis (see Woods et al., 1994).

improve performance. Some of these involved relatively simple perceptual and memory aids. For example, the electronic highlighting in Figure 4 helps the problem solver keep track of intermediate results. Other strategies involved having the computer do basic clerical housekeeping and tracking of tasks.

Since the research had uncovered expert strategies, especially ones that guard against or help recover from possible misidentifications, the designers developed a "critiquing" or advisory component in the computer system. An example of the output of this critiquing component is shown in the message in the lower window in Figure 4. In effect, the computer would "tap them on the shoulder" and say, "wait a minute, perhaps you should consider this before you go on: you just ruled out an antibody that seems inappropriate or you just left this panel without making some inferences that I (the computer) think are possible."

The researchers went on to explore different ways to use this "intelligent" capability. There are multiple ways to use this computer capability, and they are not all equally effective in improving the overall performance on the identification task. The overall performance of the human-computer team was significantly better when the human solved the identification problem and the computer provided a critique of the method used than when the computer solved the problem and the human provided the critique to ensure the computer had a correct solution. In both cases the knowledge in the computer was the same; the difference in performance was a result of the roles assigned to the people and to the computer.[7]

This conclusion is consistent with other studies, namely that team performance is better when the computer plays the role of the critic. The difference can be large, in some cases as much as 30% better. This most probably is the result of a framing effect. This refers to the way that a suggestion from the computer can limit the variety of different possibilities that the human operator explores. When the computer suggests, practitioners tend to follow this proposed reasoning and agree with the computer even when the quality of the initial assessment by the computer is poor.

The timing and character of the computer critiques are also important in determining the overall performance of the human-computer team. In the end, the researchers were able to significantly reduce misidentification rates. According to their study, students who finish their formal curriculum and then spend 2 or 3 hours in this kind of learning environment improve their performance on test cases by about 90%.

[7] Note that the cooperative system is larger than just the human-machine team. It can also include multiple people, techniques to catch misidentifications, and certification processes, among other elements.

In a laboratory setting, with a critiquing system present, performance can improve between 30% and 60%.

The research project is complete. A system is currently available as stand-alone tutoring software for the cost of media. Several labs are using the software on their own. However, there are no formal technology transfer mechanisms in place, no active assessment programs to guide the transfer, and no software maintenance or support available. All these components necessary to translate research into improvements in safety are missing.

Implications of the Research

New technology is often proposed as the solution to a "human error" problem. This project illustrates that, while technology may well be part of system improvements, technology alone is not sufficient (Woods et al., 1994, chapter 5). In this case, the critical information to decide how to use technology skillfully came from:

- understanding what constitutes hard problems,
- understanding the ways in which the task of identification is vulnerable to failure, and
- understanding the strategies experts use to guard against and recover from trouble.

Note how the research results are not mere details of implementation or user acceptance that can be dealt with after the basic concept for new technology is implemented. The studies helped discover how to use technological possibilities to aid performance. The result was a cooperative concept that was quite different from the more autonomous machine that, in the absence of good data on the nature of expertise and failure, some expected to build.[8]

The research strongly supports the use of decision-support tools to improve human performance. It also provides a warning about the limitations of automation as a replacement for human expertise. This case, with many others, provides an explanation for the repeated failures of efforts directed at replacing human expertise with machines. The results show that autonomous computer problem solvers are brittle,[9] that people's judgment can be adversely influenced by the computer's behavior, and

[8] One part of the original research team was interested in expert performance in order to develop a machine that could perform the task automatically. Interestingly, it turned out that cases that were hard for people were also hard for an artificial intelligence software system. As a result, the project emphasized using technology to support and cooperate with people.

[9] Problem solvers are brittle when they have narrow scope of competence and are unable to cope with problems that fall outside of that narrow scope. Research on autonomous machine problem solvers consistently finds that such machines are brittle. People, on the other hand, can be effective at adapting plans to handle complicating factors, surprising variations, and novel combinations. In other words, human and machine problem solvers are vulnerable to different kinds of failure.

that a carefully constructed cooperative system that coordinates both human and machine expertise performs better than either one alone.

The work demonstrates quite clearly that creating effective computer-based decision aids is itself a complex task. It requires detailed knowledge of the ways in which human expertise is deployed, how it achieves success, and how it is vulnerable. There are few applications of computer aids to decision making in medicine that have been developed with such an understanding of the cognitive demands of practice in place. Ultimately, improvement in the overall system performance depends on increasing expertise, not replacing it. The design work revolved around developing mechanisms to enhance expertise. But expertise is already highly refined in complex work domains. Improving on it requires detailed understanding of the strengths and vulnerabilities of the current knowledge and strategies.

The value of the research is not limited to the development of a specific decision support tool. The knowledge necessary to produce the computer system can be used in many other ways. Discovering the components of expertise and making them explicit permit us to consider other ways to deliver enhanced expertise where it is needed. For example, the blood banking community uses case studies for testing expertise and performance in hematology labs. The knowledge base about difficulties, typical misconceptions, and expert strategies developed for the tutor system also could be used to enhance this process.

Expertise is not simply individual skill and knowledge about the narrow technical aspects of problems; it also refers to how an organization develops, supports, deploys, and brings to bear this narrow technical expertise in different kinds of situations.

In searching out the vulnerabilities in the current system for antibody identification, the research demonstrates that success depends at least as much on effective mechanisms for *detection* and *recovery* from incipient failure as it does on the primary prevention of flaws. The success of blood banking does not arise from the elimination of errors in the antibody screening process. Rather it relies on the recognition of cases that are special and prone to failure together with indications that the identification process has gone awry. The system does not perform flawlessly at each stage but rather manages to incorporate sufficiently sophisticated detection of flaws so that the overt failure rate is very low. One component of this is the expertise applied to screening; there are others.

This critical role of detecting and recovering from incipient failure is a fundamental finding of the new look at the human contribution to safety.

This finding stands in stark contrast to the erroneous and overly simplistic notion that people are erratic and unreliable components in an otherwise successful system.

During the workshop a question was posed: if misidentification rates are so high for certain classes of problems (in cases with noisy data, weakly reacting antibodies, masking, misidentification rates approach 50% in data from practicing technologists), why aren't there more overt failures (transfusion reactions)? There are a number of reasons. The system is relatively tolerant of misidentification. The antibody identification process identifies only candidate blood units. Cross matching of the candidates with the patient's blood will detect many (but not all) incompatibilities. Even when there are incompatibilities, transfusion of the wrong unit may generate no significant problem, the problem may go unrecognized, or it may be attributed to some other source. The situation is, like so many others, complex. But an important consequence of the apparently low rate of mismatching as a source of injury is that we may expect to find little enthusiasm for expensive new programs to improve human or system performance. Paradoxically, the typical success of the blood testing system and the low-frequency of the sorts of complex identification problems used as test cases in this research may lead many to discount the value of such research.

Uncelebrated Case #3: Drug misadministrations via computerized infusion devices in the operating room[10]

Cook and colleagues studied how the characteristics of a particular infusion device used in cardiac anesthesia contributed to a series of operating room near misses. In this uncelebrated case, the story of how the incidents were investigated and how different stakeholders reacted is as revealing as the specific results.

The research began after an institution experienced an inadvertent delivery of a vasoactive drug via a computerized infusion device during cardiac anesthesia. Due to prompt physician intervention, the misadministration had no lasting consequences for the patient.

The researchers, who were already engaged in a study of human performance in anesthesia, began to investigate the incident in particular and to study broader questions about physician-device interaction. In the midst of these studies three more misadministrations occurred (again with no lasting consequences).

These studies of device use in context and the incident investigations showed that the device possessed classic deficiencies in human-computer

[10] See Cook, Woods and Howie (1992), Moll van Charante et al. (1993), Yue, Woods, and Cook (1992).

interface (HCI) design.[11] These HCI deficiencies contributed to misoperation and misassembly of the device. These HCI deficiencies made it difficult for users to detect and recover from these and other problems. These HCI deficiencies were one contributor to incidents of misadministrations of vasoactive drugs. The results also led the research team to design an alternative device interface and displays to illustrate how to correct these kinds of HCI deficiencies in this class of infusion devices. The study has implications for incident reporting, for device design, and for the analysis of human performance in technical environments.

Background

Infusion devices are ubiquitous in medicine, as are problems related to their use. The incidents with this infusion device occurred during anesthesia for cardiac surgery. Drugs with rapid onset and short duration of action have the advantage of permitting quick adjustment (titration) to achieve desired effects. In cardiac surgery there are predictable periods where patients may require the infusion of these fast-acting/short-lasting drugs to increase cardiac contractility, change blood pressure, or alter heart rate. Various mechanical devices are used to administer these infusions. The advent of microprocessor-based, battery-powered infusion devices opened the way to more precise control of these infusions than was possible with older, purely mechanical devices. But this increased precision has been achieved at the cost of increasing the complexity of the drug delivery process and the creation of new forms of failure.

The Research

Table 2 presents the sequence of events that followed the first case of inadvertent drug delivery.

The researchers were engaged in a study of human performance in anesthesia. The first incident was reported informally to one of the investigators shortly after it occurred. The initial descriptions were vague, but the event involved free flow of a vasoactive drug through an infusion device. Free flow is a runaway condition where the fluid containing the drug is delivered to the patient as a continuous, unlimited flow rather than as a controlled incremental delivery over time. In this instance, the fluid contained a drug that lowered blood pressure. Other drugs were given to counteract the effect, but the free-flow event was not discovered until later.

At the time, the failure was attributed to human operator error. This was the view of the manufacturer and also of the senior practitioners. For

[11] These deficiencies in human-computer cooperation are called classic because they have been identified as contributors to erroneous actions and assessments in many different settings. Because they are common design errors in computerized devices, they are used as cautionary tales when teaching HCI.

Table 2. The sequence of events in the investigation of four operating room incidents involving misadministrations via an infusion device

Background: *Investigating incidents related to human performance in anesthesia*	• case protocols collected prospectively based on cases presented at the internal morbidity and mortality conference
Investigation I:	• informal notification of an OR incident involving infusion device (no patient consequences) • interviews with participants within hours of incident to reconstruct case • bench testing of device behavior to corroborate sequence of events and to identify underlying contributors • results identified classic deficiencies in practitioner-computer cooperation
Reactions to incident by practitioners and management:	• human error— "can't make devices idiotproof" • coping strategies—"yeah, there are some device weaknesses, but I can handle it" • untrustworthy device—"you have to be careful, it can burn you"
Investigation II:	• exactly one week later a second incident occurs: another near miss • device captured in "impossible" state • interviews with participants immediately following incident to reconstruct case • investigation identifies central role of breakdowns in practitioner-computer cooperation • other methods employed to understand device use and breakdowns in context: observations of device use in context; more testing of device behavior • unable to get any useful data on other incidents involving this device or similar devices from incident reporting systems
Reactions by stakeholders:	• cryptic incident report by device manufacturer refers to "custom" setup, states device worked as designed, implies erratic human behavior was responsible • device manufacturer reaction focused on, is there a patient injury? Will we be sued? • Practitioners now see device interface as the source of incidents and difficulties; they report more cases
Investigation III:	• two more incidents are reported and investigated "fresh" • report of the results of all of the investigations and studies documents the problems in practitioner-computer cooperation and how they contributed to incidents (Moll van Charante et al., 1993) • results lead to predictions of other problems; later, incidents occur where these problems are one contributor to the sequence of events (e.g., an event during transport to ICU) • researchers begin a follow-up project to redesign the device interface; the goal of the redesign is to show how to correct deficiencies in practitioner-computer cooperation (a) with this device, (b) with this class of devices and, (c) in general (Yue, Woods and Cook, 1992)
Aftermath:	• investigators report incidents and results in specialty journal (Cook, Woods and Howie, 1992); give talks to research-oriented and technology-oriented anesthesiologists • use of device is reduced • leads to studies of different classes of infusion devices used in other contexts (Obradovich and Woods, 1996)

some, the event was regarded as an example of human frailty, or at least the limited ability of humans to operate modern equipment (e.g., "you can't make devices completely idiotproof"). Many practitioners thought that the device had some quirks in operation and that it was potentially troublesome. Some had developed local adaptations to help them forestall problems with the device in use.[12] In general, other practitioners felt that this kind of thing could not happen to them because of their skill, attention to detail, and vigilance. In addition, the event was regarded as an anomaly with only local implications, unrelated to other events.

To the researchers, however, the incident had the flavor of a human-computer interaction breakdown. They were familiar with problems in human-computer cooperation and also with investigating human performance in incidents and were already active in this setting in a related study. As a result, they decided to investigate the incident and the device more closely.

The researchers used multiple methods to reconstruct the sequence of events and to explore what factors contributed to the incident. They explored how the device behaved under different circumstances, e.g., how the alarms and displays behaved when flow was obstructed or excessive. They began observing how people used the devices in the context of cardiac surgery. They linked aspects of cardiac anesthesia to device characteristics and the user interface—for example, the need to use multiple devices in parallel in this setting. The data were used to construct a protocol of the incident that consisted of what cues the anesthesia team noticed about the patient's physiology, their interpretation of the situation, and their interventions. In particular, the protocol traced the interaction with the set of infusion devices during the case.

The basic sequence of events was as follows. The anesthetist observed increasing blood pressure and attempted to counteract the change by starting one of the infusion devices that had been set up earlier to be ready to deliver medication to lower the blood pressure. The device emitted an audible alarm and posted a message on its screen indicating that no flow had occurred. An anesthesiologist scanned the assembly and noted that all of the stop cocks were closed downstream of the infusion devices, blocking flow to the patient. The anesthesiologist then opened all

[12] In one case, a practitioner was observed setting up the devices. Setup was completed at the beginning of the day, well before bringing the patient into the room. After assembling the devices and drugs the practitioner started all the devices at a high flow rate setting, allowing the fluid they controlled to flow into a garbage pail. Once the devices had been running for several minutes without generating any alarms, the practitioner changed the settings to low rates and shut the devices off. This pretest of the assembled system of devices is a good example of local adaptation.

of these valves. By this point in time blood pressure had fallen. Because the infusion was no longer needed, the anesthesiologist did not restart the device (in the anesthesiologist's mind the infusion had never started).

The blood pressure began to fall and reached an unacceptably low level. The anesthesiologist responded appropriately by injecting other drugs to counteract the drop. However, the disturbance to blood pressure continued, and the anesthesiologist continued to act to keep blood pressure under control. When the anesthesia team scanned the array of infusion devices, the displays indicated that the one they originally attempted to use was not running. Even so, they pressed the OFF button on the device to turn off the power. Only later, as they began to prepare another drug infusion to counter the low blood pressure, did they notice that the previously full bag of drug for lowering blood pressure was now empty.

Although they did not realize it at the time, the anesthesiologists had misassembled the device in a way that allowed free flow. The closed stop-cock in series prevented the immediate free-flow condition in the infusion device. When the anesthesiologist opened these valves, free flow began, drug reached the patient, and the blood pressure began to fall. Once the unintended drug delivery began, the device provided no feedback to indicate flow of any kind to the users. The display and alarms indicated there was no flow and that there had been no flow. The device's sensor obscured the user's view of the drip chamber. The bag of fluid that contained the drug was inside a veil of aluminum foil to prevent it from reacting with light, thus obscuring the user's view. Furthermore, the off button only powered down the device; it did not block flow. Figure 5 contains the final version of the protocol describing the incident.

The studies of user-device interaction in context showed how multiple factors could come together to produce this incident. Some of the factors were traps created by the device interface and displays. For example, because several of these devices were used together in a device array and the setup for each individual device was complex, there were several steps that might be omitted or incorrectly performed which could produce a path toward failure. Misassemblies were observed to occur during normal use, but they did not produce inadvertent drug deliveries because other necessary conditions were not present. The observations showed that users were sensitive to the possibilities for misassembly and misoperation and devised strategies they thought would help to avoid these or to prevent inadvertent drug flow the patient. In the context of cardiac anesthesia, a misassembly (i.e., a flaw in setting up the device) could occur much earlier than the effects of the failure (i.e., the moment that drug began to flow freely).

If an unintended drug delivery began, there were other characteristics of the device that made it difficult for operators to detect that something had gone wrong with one device in the array and to correct the failure. Basically, the device provides weak feedback about its activities. Under the right circumstances the device's alarms and displays can give the impression that there is no flow when in fact there is flow or the impression that flow is precisely as desired when in fact it is different.

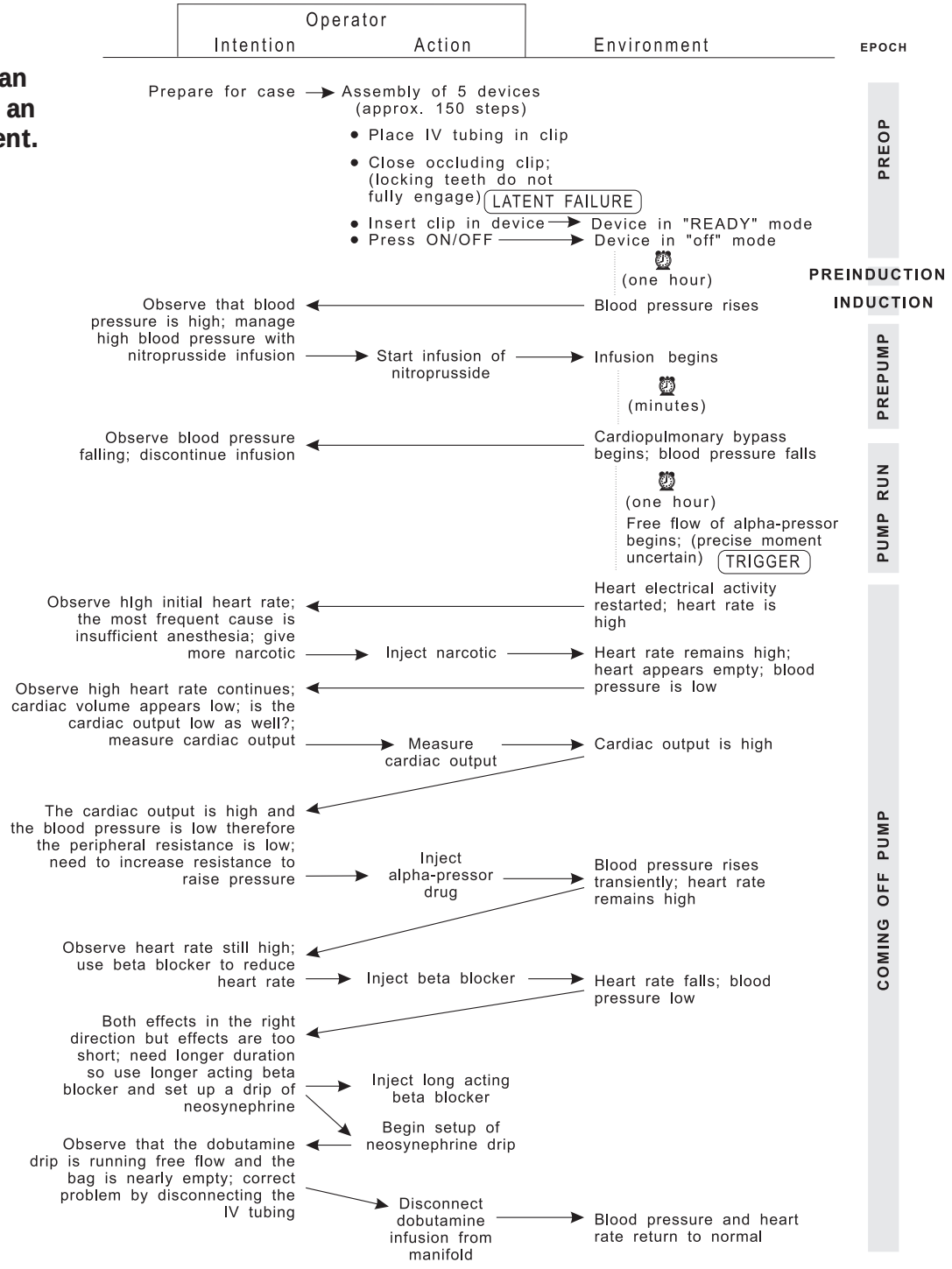
Up to this point the failure was regarded as a operator error, or perhaps a training problem, but only of passing significance. One week later, however, another near miss occurred involving this kind of infusion device, with one of the senior practitioners using the device. The anesthesiologists recognized that the surprising cardiovascular behavior resulted from the behavior of the device. The device was in a state that all thought was impossible—the device was delivering drug even though it appeared to be "off" as evidenced by a completely blank screen. The research team was called, and they removed the device to a place where it could be studied, thus capturing the device in its failed mode.

With a video recorder running and the manufacturer's representative present, the investigators determined the specific contributors that created the impossible state. This was done by varying the setup and operation of another of these infusion devices until its behavior matched the "impossible" behavior of the operating room (OR) device. The results from the previous investigations of device use in context and device behavior were important contributors to this process.

At this stage, the research project began in earnest with detailed, formal field studies observing the user setting up, testing, and using the devices. To verify how the device actually behaved and how it appeared to behave under different normal and abnormal conditions, the researchers tested the device in an engineering laboratory under a variety of conditions. The user interaction sequence was worked out for various tasks to reveal various HCI problems (e.g., ambiguous alarms and multiple hidden modes of operation). These results made the problems with the device in clinical settings comprehensible. They also made it possible to see which aspects of the user-device interface created opportunities for problems in the context of cardiac anesthesia.

In the research the investigators were unable to get any useful data about other incidents involving this device or similar devices from incident reporting systems. The device manufacturer provided a brief, cryptic incident report to the FDA device incident reporting system which referred to a "custom" setup, stated that the device worked as designed, and implied erratic human behavior was responsible. But practitioners

Figure 5. Protocol describing the interaction between anesthesiologists and an infusion device during an operating room incident.



now saw the device itself as the source of difficulties or surprises they experienced, and they began to report more incidents involving the device.

During the research projects, two more incidents related to the device occurred and were investigated. Several new reports of difficulty with the device were also received during this period. Some of these helped to confirm findings about difficulties with the device-user interface.

The research team also began a follow-up project to redesign the device interface (Yue, Woods and Cook, 1992). The goal of the redesign was to show how to correct deficiencies in practitioner-computer cooperation (a) with this device, (b) with this class of devices and, (c) in general. The design project used the results on the HCI problems to create a redesign based on user-centered automation principles. Particular attention was paid to:

- making the device display its actual and intended functions in a way that allowed users to see whether the actual performance of the device matched the intended one;
- supporting the user's need to attend to other tasks and interact with the device only at intervals by making the device show its behavior over time;
- making it possible to switch smoothly between automated and manual methods of control, so that situations where using the automated system would lead to instability (like transporting patients connected to the device from one location to the other) could be handled by taking manual control;
- providing a direct, positive, visible control that stopped flow through the device; and
- incorporating specific features to reduce the difficulties associated with using multiple devices simultaneously, including a support tree that made it possible to align infusion devices and their source bags of fluids and a slender package that permitted side-by-side arrangement of devices.

Broadly speaking, these features were all directed towards making the operations of the device apparent to the operator or, in the jargon of HCI, more "visible."

The results of the studies led to predictions of other problems one might expect. For example, the results alert us to the potential for certain kinds of problems and incidents in total intravenous anesthesia, which requires the use of arrays of automated infusion devices. These potential problems could be avoided largely through improved interface design. In another example, the studies showed that during transport from the OR to the ICU actual device performance was irregular in a way that was unpredictable

and invisible to the user.[13] Later, the investigators were present for discussion of a case in a morbidity and mortality conference involving a patient whom became unstable while being transported from the operating room to the intensive care unit. One of these infusion devices was being used to support the patient's cardiovascular system. While it was impossible to reconstruct the device's behavior during this period and to determine its contribution to the incident, the researchers were able to alert the physicians that the device could behave erratically and unpredictably under these sorts of circumstances.

The investigators reported the incidents and results of their investigations in a specialty journal and through presentations to research-oriented and technology-oriented anesthesiologist groups. A report was prepared that describes the results of all of the investigations, documents the problems in practitioner-computer cooperation, and traces how they contributed to incidents (Moll van Charante et al., Cook, Woods, Yue and Howie, 1993). Overall, the research on the device lasted nearly nine months. It was not funded; the participants donated the time for the project.[14] The work also led to other studies of different classes of infusion devices used in other contexts (e.g., Obradovich and Woods, 1996).

Implications of the Research

One interesting feature of the research is that the deficiencies of the device design are subtle and became apparent only under the conditions of use. Problems occur when aspects of the context of use combine with features of the device and device interface to create problems for the user. The testing that uncovered specific problems with the device-user interface was directed by the field studies that in turn were prompted by close examination of incidents. The device did not possess a hidden "Achilles' heel" defect but rather possessed a group of properties that influenced human performance. These factors were significant with respect to outcome only under specific circumstances.

It is important to note that the incidents were regarded as operator error until the investigation was well underway. A variety of factors tended to make it unlikely people would discover the specific problems with the device in this context of use: a) the complexity of the larger system in which the device was used; b) the hidden complexity of the device itself; and c) the general experience of practitioners that computerized devices are quirky, difficult, or unpredictable.

13 The device monitored its own function by counting fluid drops forming in a drip chamber, but lateral acceleration could move these drops out of the detector's path; this could lead to a variety of responses (e.g., automatically blocking flow and then trying to resume flow at the target setpoint) depending on precisely how and when the detection failed along with other factors.

14 Moll van Charante was a visiting medical student from The Netherlands, Yue was a graduate student in the Industrial Design program focusing on human-computer interaction at The Ohio State University.

These factors made human performance rather than device characteristics the center of attention after the fact. Indeed, the reports regarding incidents with this device that were filed with the government incident tracking system emphasized simple "operator error" and implied each was unique (e.g., "user reports problem; cannot duplicate problem"). None of the reports provided a narrative of the incident with enough detail for researchers to go back and look for similarities or contrasts with new cases. None of the reports indicated any in-depth investigation of the factors that led to the incident.

Only when these actual incidents were carefully explored, applying specialized knowledge and techniques related to the factors that affect human performance, were investigators able to reveal a second story hidden behind the label of "operator error."

Another important feature of the research is the use of multiple methods to understand the different factors at work. Each new finding based on a particular method raised questions that required a shift to a different method. This is natural, considering the complexity of the underlying features of the domain. The ability to make progress depended on being able to bring together disparate methods to create a web of information that was mutually reinforcing. The incidents themselves pointed to features of the device. The HCI analysis of the device suggested particular problems with the interface. The studies of the behavior of the interface showed how the device would appear opaque under conditions like those occurring in the incidents. The redesign showed how this opacity was a function of cognitive tasks and how it might be avoided without changing the underlying mechanical functions of the device. The contrasts and connections between these various approaches provided the insight.

Finally, the research is significant because it was so fortuitous and unplanned. It points out the value of long term associations between researchers and practitioners. In particular, the ability to recognize fruitful areas for investigation depends on being intimately involved with practitioners. The research perspective allows one to see beyond the practitioners' own characterizations of the difficulties they face and to follow deeper, more subtle, but ultimately more rewarding lines of inquiry.

Contrasting Uncelebrated and Celebrated Cases

Taken together, the laparoscopic cholecystectomy, blood antibody identification, and infusion device cases demonstrate the kinds of insights that come from exploring the second story that lies behind the incidents that provoke attention. In each case, the work is painstaking and

detailed, going far beyond the sorts of investigations that followed the celebrated cases. In each case the story is complex, difficult for outsiders to understand, and not easily reduced to a simple summary. Significantly, the research methods used are unfamiliar to many. Finally, the motivation for the work was less the desire to directly generate safety improvements than to understand the nature of the real processes that underlie success and failure in the real world. The potential for such work to produce sustained increases in safety is substantial. In particular, in each case the research offers the possibility of further progress by identifying areas ripe for additional work.

Day Two – Incident Reporting and Analysis

The issues surrounding incident reporting systems were the focus of discussions that began the second day of the workshop. The history of systematic incident reporting and analysis in medicine is a rich one, extending back at least to the targeted efforts of Cooper et al. (1978) to generate and analyze patterns in a corpus of cases in anesthesiology. Today, there are a number of such systems in place in health care and a variety proposed or in development.

The interest in incident systems is spurred by several different beliefs:

1) the belief that there exist a variety of patterns in the character and occurrence of incidents that go unnoticed because there are no larger, continuously replenished, systematically generated collections of data;

2) the belief that the analysis of these patterns can be used to direct attention to the areas most rewarding for study and amenable to improvement;

3) the belief that the present pace and character of technological, organizational, and economic change in health care is shifting the pattern of incidents; and

4) the belief that the absence of data defining these patterns will prove to be the critical, limiting factor in improving safety.

Closely linked to these beliefs are experiences with existing incident reporting systems. While there is no real method for measuring the performance of existing systems,[15] the view is widespread that less than 5% and perhaps less than 1% of incidents that might fit the criteria for reporting are actually reported. The existing systems are mainly mandatory, and many are linked either directly or indirectly to enforcement and sanction mechanisms.

[15] The reason for this is that knowing the rate of reporting requires knowing the denominators for numbers of events; that is, knowing precisely what it is that the incident reporting system is supposed to be discovering.

Many leaders in health care feel that new approaches to incident reporting are required. However, most incident reporting discussions revolve around how to achieve greater compliance with reporting requirements. Proposals for anonymous systems, confidential systems, immunized systems, or mandatory systems are framed primarily by concerns for gaining more (greater numbers, more detailed) reports.

The discussion during the workshop explored incident reporting in health care from different perspectives. The stimulus for the discussion was several short presentations on lessons learned about incident reporting and incident analysis from other industries. The presentations generated a discussion that focused more on how the analysis of reports is complicated, difficult, and sometimes controversial. The discussion was wide-ranging and complicated. Topics included:

- building consensus among stakeholder groups,
- analysis of incidents with respect to factors influencing human performance,
- complexities and limits in the attribution of "cause,"
- linkages between incidents and accidents, particularly in health care,[16]
- difficulties in using incident data to improve safety.

The session opened with a short talk by Charles Billings, MD, Chief Scientist (retired), NASA Ames, on the lessons learned from incident reporting in aviation. Dr. Billings designed, started and managed the Aviation Safety Reporting System (ASRS) 22 years ago when he was at NASA's Ames Research Center.

The ASRS is a confidential reporting system for incidents and not for accidents. It is often proposed as a model for incident reporting in health care. Dr. Billings described the history of that system and the conditions that now appear to have been critical for its success (interest in the aviation experience is widespread in medicine; as a result, Appendix B contains an edited transcript of Billing's presentation.)

Lessons from the Aviation Safety Reporting System (ASRS)

The ASRS is operated by NASA and largely funded by the Federal Aviation Administration (FAA). It is a successful system that was developed in part because of the failure of a predecessor system run from within the FAA. Because the FAA is a regulatory and enforcement body, reports to that system were limited. The ASRS was developed as an independent system, run entirely outside the FAA, and was, from the outset, designed to be entirely confidential. Reports made to ASRS include an identification strip that provides analysts the means to contact the

[16] In aviation, there is a reasonably clear demarcation between categories labelled as "incidents" and "accidents." "Accidents" is used to refer to cases where passengers are injured or where there is overt damage to the aircraft. The term "incidents" refers to cases that violated some aspect of good practice or rules but did not lead to injuries. Despite these working definitions in aviation and other fields, the links between good practice and outcome are complex. In medicine the links between good practice and outcome are even more difficult to untangle.

reporter. This strip, and anything that would uniquely identify any individual, is removed during the analysis. The narrative description of the incident is retained as are a host of indexing keys. Incidents are collected and reported to the aviation community as individual episodes and as exemplars of larger problems. The larger database of incidents is available for research. Each year there are on the order of 30,000 incidents reported. The system costs several million dollars per year to run.

A consensus among stakeholders that such a system is needed was essential to the continued success of the ASRS. Producing this consensus was a substantial effort in itself. Some portion of the success of the ASRS was derived simply from creating the consensus. The effort needed to acquire agreement among the stakeholders created an environment that nurtured the system and protected it from political tampering when its output was controversial. But creating the consensus also generated a widespread (but not universal) view of safety that insisted that practitioners (pilots, air traffic controllers, mechanics, flight attendants, etc.) were the observers most likely to recognize hazards and incidents and were also vital in preventing bad outcomes. The goal of collecting the details surrounding "accidents that might have happened" is to identify previously unknown hazards and to see new emerging threats as systems and organizations change. The system generates this type of information by performing analyses of sets of narratives as questions about threats to safety emerge. It does not generate large statistical measures of systemic performance—a fact that was stressed repeatedly.

The analysis of incidents reported to the ASRS depends on a cadre of analysts with multiple skills. These individuals are domain experts (e.g., pilots) rather than technicians or clerks. The point was made several times that the analysis of the reports requires at least as much expertise as is involved in their generation. Researchers can also make use of the database by working with the staff of analysts to put together subsets of narratives that address a particular theme or question. The analysis also depends on the ability to contact reporters to clarify details of the incident. These activities depend on an effective indexing scheme so that analysts can put together related or contrasting sets of cases for analysis. Note that, although the system uses substantial indexing, the primary purpose of analysis is not to reduce the incident to a category but rather to make sure that the narrative is descriptive, complete, and precise. Because the ASRS is not fundamentally a statistical system, the substance of the narratives is the critical information that the system provides.

A critical part of the activities of the staff at the ASRS is providing feedback to the operational community—the people who voluntarily provide the information. The staff uses several mechanisms such as the *Callback* newsletter to provide highly visible, monthly feedback to the community of the results of its analyses and studies of the data received.

The visibility of the information provided by individual reporters back to the operational communities has proven to be an essential part of system success, building support for the system and making safety a tangible value.

The ASRS does not provide guidance about how to solve problems or about which problems are economically or socially worth attention. It has no regulatory function. It does not deal with accidents, which are reported and analyzed separately through the independent National Transportation Safety Board (NTSB). Studies using the ASRS data base have been motivated by accidents and have proven helpful to the NTSB in understanding the contributors to an accident it is investigating.

Reports to the ASRS for specific incidents provide limited immunity against FAA enforcement action but only under specific circumstances. This immunization of the reporter has itself been an incentive to report and has led to a substantial continuing flow of reports. Technical developments in the aviation system have allowed for automated detection of "altitude busts" where an aircraft strays outside its assigned altitude. This has created an incentive for pilots to report such incidents to the ASRS in order to be able to claim immunity against later disciplinary action. Viewed from one perspective, these reports are monotonous and repetitious. They are, however, more informative than the automated detection system, which simply records the event. The narrative descriptions can provide information about how and why such "altitude busts" occur. Such information has provided the basis for procedural modifications designed to ameliorate the problem in several air carriers. Nevertheless, it is clear that the incentive of immunity affects the number and kind of reports received.

There have been no breaches in the confidentiality of the ASRS system. Narratives entering the database are "de-identified" in a process that removes all the features of the report that might be used to identify the event and people it describes. This process takes priority in handling ASRS data. It provides effective immunity by transforming the data into a form useless for civil sanctions. It is clear that the reputation of the ASRS among practitioners is derived in large part from the record of success in providing such functional anonymity.

The impact of the ASRS on safety is partly indirect. Simply by its presence it has served as a potent indication to all the stakeholders that safety is a critical concern, that new hazards will continue to appear, and that there is a system-wide concern for safety that arches over all organizational and institutional boundaries.

The above lessons are abstracted from the aviation experience. Both in the presentation and the ensuing discussion, the workshop explored important differences between health care and aviation. While a

successful system for aviation is not likely to transfer directly and literally to health care, the lessons Dr. Billings has derived are generic, e.g., a non-punitive approach, the importance of communication back to practitioners, and the critical role of an independent organization. As such, these lessons can serve as a guide to develop successful systems in health care.

Incident Classification and Analysis

Collections of incidents and accidents cry out for classification. The apparent similarities and differences between the events, their outcomes, and the circumstances that precede them encourage us to organize them in categories and rank them in severity. But classification also has its own hazards, especially in complex domains where there are multiple possible paths to any outcome and multiple possible outcomes from any path. Classification involves identifying relevant similarities and differences; their effective use depends on being able to know *a priori* what *relevant* means. Erik Hollnagel, an expert in the evaluation of human performance, explained some of his experience with classification systems used in industrial incident and accident work (see Hollnagel, 1993). His examination of these sorts of systems revealed that an extensive effort at *a priori* classification may yield very little insight into the underlying features that incidents have in common.

In the discussion about incident reporting, it was pointed out that the ASRS uses an extensive indexing system, but this is used to collect related subsets of narrative cases from the database that pertain to a theme or question. The indexing system does not work automatically but is a tool used by the staff to carryout analyses and to assist outside parties use the database in their analyses. The indexing is used as a tool in analysis; the classification system it represents is not the analysis.

Classification does involve a type of analysis but a type that greatly constrains the insights that can be obtained from the data. Typically, when classification systems are used as the analysis, a report of an incident is assigned, through a procedure or set of criteria, into one or another fixed category. The category set is thought to capture or exhaust all of the relevant aspects of failures. Once the report is classified the narrative is lost or downplayed. Instead, tabulations are built up and put into statistical comparisons. Put simply, once assigned to a single category, one event is precisely, and indistinguishably like all the others in that category.

Yet research on human performance in incidents and accidents emphasizes the diversity of issues and interconnections (e.g., Woods et al., 1994). As Billings emphasized in the discussion of the ASRS, capturing a rich narrative of the sequence and factors involved in the case has proven essential. Often, new knowledge or changing conditions leads investigators to ask new questions of the database of narratives. The analyst often goes back to the narrative level to look for new patterns or connections.

As an example, Hollnagel described an industrial incident reporting system that in one sense seemed a success but in another sense failed. It was successful in that people reported to the system, but it was a failure in that these reports did not lead to significant learning about vulnerabilities or to constructive changes. The central reason for this failure was the removal of the interesting, informative aspects of the events that were present in the narratives but lost in the process of classification.

Hollnagel traced the failure, in part, to the classification system's failure to distinguish between the phenomenal appearance of a failure event and the underlying pattern of contributing factors that generated the event. To use a medical metaphor that Hollnagel has employed, most classification systems confuse phenotype with genotype. The phenotype of an incident is what happens, what people actually do or what they do wrong, what you can observe. Phenotypes are specific to the local situation and context—the surface appearance of the incident. On the other hand the genotype of an incident is the characteristic collection of factors that lead to the surface, phenotypical appearance of the event. Genotypes refer to patterns of contributing factors. The significance of a genotype is that it identifies deeper characteristics that many superficially different phenotypes have in common.

Genotypical patterns are not observable directly. All statements about them are inferences that represent models about the factors that drive human performance rather than observations. It is simple to state the difference between these but quite difficult to separate them in practice. What reporting systems provide are phenotypes. What drives performance, however, are genotypes. The processes of inference about the contributors to events depend on a thorough understanding of the background or context of the event. The uncelebrated, researched cases illustrate the process of finding possible genotypical patterns. They also illustrate how finding these patterns can help identify meaningful positive interventions to enhance safety.

Incident collections do spur interest, in part because of the contrasts and similarities between cases. But classification systems that rely on phenotypical categories do not capture these characteristics very well. Indeed, many at the workshop noted that "human error" is nearly always an important category in classification systems for accidents, but assigning a case to this category generally stops or limits the analysis of what factors influenced human performance.

Classification systems that obscure, simplify, or discard the story of the cases they classify have generally not been successful. The systems themselves become outdated relatively quickly. More significantly, the collections they represent generally lead to little real progress on understanding the nature of success and failure in complex domains. Even when motivation is high in management and there are high

consequences of failure, the process of classifying by phenotypes eliminates the ability to see the second story of contributors to the system failures. Classification systems limit the depth of the analysis that can be conducted, and they limit what it is that one can learn from the collection of data. This is especially a problem in complex environments where failures do not occur because of single causes (Reason, 1990; 1997). The net result is that classification systems tend to strip away the rich contextual information from which inferences about genotypes may be made and thereby make such collections sterile and uninformative.

Although their methodologies differ, virtually all the researchers present at the workshop commented that their work depends on capturing the process and the context that led up to the outcome. This "story" is the fundamental data, and all analyses build up patterns, trends and contrasts across these stories. From a research perspective the sparse, simplistic stories of the celebrated cases were not so much wrong as they were uninformative; the researchers did not see a way to make progress based on those kinds of data. Rather, it was the richer stories that captured attention and served as examples in the conversations during the workshop.

Incident reporting is one way to obtain such rich stories. But this method of gathering data is largely passive. There is no way to obtain data other than by encouraging practitioners to send back reports when things go awry. Other, more active approaches are also possible.[17] Gary Klein has conducted many critical incident studies to better understand the nature of expertise in complex settings (see Klein, 1998), and he commented on other approaches that can be used to generate collections of incidents.

In Klein's technique, researchers first proactively go out to practitioners and help them recall and walk through past incidents. The focus of these discussions is to help practitioners generate cases that illustrate the nature of expertise, show how they succeed, demonstrate what makes problems hard, and reveal how failure occurs. As in the uncelebrated cases, contrasting success and failure provides critical insights. The analysis is an involved process that extracts the critical factors in the story and shows the interplay between these factors. It depends on concepts and models about the factors that affect human performance (genotypical patterns). It looks for patterns and contrasts across a set of cases that speak to an issue or question.

The studies in the uncelebrated cases illustrate this kind of active research process. They illustrate how the results provide insight about how the system works much of the time but how it is also vulnerable to failure. They illustrate how this insight can guide investments that will enhance safety. The uncelebrated cases are not simply specific places where this

[17] An early example in medicine of the use of these active techniques is a series of studies by Cooper and colleagues in anesthesia in the late 70's (e.g., Cooper et al., 1978). Based in part on the classic work in Human Factors (Flanagan JC. The critical incident technique. *Psychol Bull.* 1954; 51: 327-358.) Cooper used proactive techniques to better understand the landscape of safety in anesthesia.

learning has gone on, places where we are ready for the work to develop and test enhancements. They are also markers and beacons for the kind of process that is needed to better understand the vulnerabilities in other areas of health care and to see new ways forward to enhance safety.

Learning from Incidents and Accidents

The discussion at the workshop considered the many issues associated with analyzing incidents or accidents—how we learn from such events.

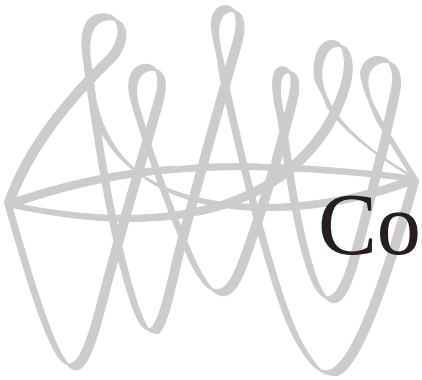
As failure rates fall, the ability to learn individually or collectively from failure falls as well. The meaning of a particular failure will be ambiguous and contentious. The multiple contributors each necessary but only jointly sufficient for the accident, complicate the ways in which accidents are investigated and understood. This makes the attributions of cause complex. In turn, these characteristics of the post-accident aftermath influence the learning process in several ways. At least two are worth mentioning here.

First, many accident investigations end prematurely. After the fact, people only see the ways that practitioners at the sharp end could have acted differently because of hindsight bias. The variety of organizational and institutional factors that influence the decisions and actions at the sharp end are unexamined or discounted. The risk of ending the investigation early is great. Taken as a whole, the research studies show that organizational factors play a critical role in fostering events and create vulnerabilities and latent failures that contribute to events.

Second, failure is often seen as a unique event, an anomaly without wider meaning for the domain in question. Post-accident commentary typically emphasizes how the circumstances of the accident were unusual and do not have parallels for other people, other groups, other organizations, other technological systems.

The narrow focus on human error as the cause of the accident serves to reinforce this view. If a given accident is caused by isolated human error then the accident is without deeper meaning. After all, the reasoning goes, the human performance in the accident was so egregious that it cannot possibly have meaning for us here. We are more careful. We are more conscientious.

Emphasizing differences blocks the learning process. High reliability organizations appear to recognize that incidents mark vulnerabilities and threats that could indeed happen to them (see Weick and Roberts, 1993). They search for levels of analysis that demonstrate, not the differences, but the similarities between the accident situation and others in order to find new ways to improve the larger system.



Conclusions

The conversation that took place over two days at the workshop was a wide ranging discussion about safety, accidents, and research in health care. The discussion was mainly about contrasts:

- between celebrated and uncelebrated cases;
- between success and failure; and
- between naïve attributions of failure to human error and detailed investigations of the strengths and weaknesses of expert human performance in context.

This report presents one synthesis of the materials from the workshop—it follows a few threads through the two days of talk.

Just like any real conversation, there were sometimes several topics discussed at once. Many issues were raised, questions put, and subjects left hanging without conclusion. The workshop did not attempt to develop a consensus. Rather it was an effort to obtain a variety of perspectives, to engage in an exploration of safety in new ways.

The researchers offered no solutions, nor did they identify easy paths to success. Rather they pointed out how myopic our present approaches to safety actually are. In polite and sometimes not so polite terms they indicated that fascination with the celebrated cases of failure is unlikely to yield any real progress towards safety. They encouraged research into the basis for success as a means for understanding failure. They pointed out how careful examination of seemingly peripheral questions about how practitioners work provided the new insights. They showed how research on understanding of the real tasks of real practitioners can lead to new technology that actually improves performance.

The researchers warned against narrow focus on practitioners at the sharp end, pointing out that the lessons from other industries are that accidents reflect systemic factors and not individual ones. They warned, too, against trying to treat safety in isolation from the other aspects of health care.

Rather, safety is an embedded feature of a complex, dynamic system. They were optimistic and encouraging about prospects for research that bears on safety in health care but, as seasoned researchers with long experience on these topics from other industries, they were also cautious about the recent flurry of interest in safety as a goal. Several times the conversation turned from potential for progress to warnings against efforts to improve safety directly by programs that look attractive but are disconnected to the larger research base on human and system performance. Their experience with other industries indicated that the need for the appearance of a commitment to safety can sometimes take precedence over the long, painstaking efforts required to make real progress. It is much easier to talk about a "safety culture" than it is to create one.

Charles Dickens' book *A Tale of Two Cities* begins with a famous litany contrasting the time as both the best and the worst. The title of this report is meant to evoke that same sense of contrast.

At a superficial level—at the level of the celebrated cases as they are usually presented—the story of safety in health care is about repeated, unrelated, isolated, incomprehensible accidents that dog the heels of the vaunted successes of modern technological health care.

At a deeper level, the story is about the ways in which success and failure are derived from the same sources. It is about the ways in which exposure to hazards is indivisibly connected to the pursuit of success. In this contrasting view, the bad events are not separate phenomena that can be eliminated by the use of some managerial or technological tool. Safety is not a separate entity that can be manipulated or achieved in isolation. Rather it is an emergent property of the ways in which the technical, individual, organizational, regulatory, and economic factors of health care join together to create the settings in which events—the best ones and the worst ones—occur.

Although the researchers had much to say about safety, none of them were researchers on safety in itself. Their research is primarily about human performance, technology, organizational behavior, and even philosophy. These are all fields that bear on safety and describe the ways in which the factors interact to cause safety to emerge. John Flach spoke for many of the researchers when he said, "researchers pursue interesting questions." His point was that this work may lead to new views and applications that advance safety, but that the really effective research does not start out that way. His statement can be taken as a warning about the need to establish and sustain a variety of lines of research on human and system performance in health care in order to make progress on safety. The uncelebrated cases all involved long-term efforts focused on apparently small questions. This much is clear—gaining more insight requires sustained, detailed efforts.

The situation confronting those who want to increase safety today is not unlike that confronting those who wished to eradicate cancer a generation ago or more recently those who wished to find a cure for AIDS. At first glance, these were simply applied problems that needed bigger, better, more powerful treatments of the sort that were already being applied. But the real improvement in treatment of these diseases came not from direct applications. The real improvements came from study of the mechanisms of disease, often in areas that appeared only superficially related to the problem at hand. The discovery of genetic causes of cancer, the development of the protease inhibitors that now offer HIV-infected people a chance for long life, came out of efforts that looked scientifically into the complex mechanisms that underlie these diseases. It is ironic that the research pathways that led to these successes do not immediately suggest to us similar approaches to learning about safety by studying the complex mechanisms that lead to success and failure.

The organizers of the conference thought to challenge the assembled researchers to describe how their research could be used to gain new insights into safety in health care. The researchers provided a host of pointers and engaged in a wide-ranging conversation about the opportunities and obstacles to research on safety. Several contacts between individual researchers and potential users of their work happened at the workshop. The transfusion medicine work was particularly interesting to some of the workshop sponsors. The lessons about incident reporting and analysis captured the attention of other people and organizations. But in the end, the researchers challenged the organizers and by extension, the health care community to take a new look at safety, to change long-held views about the sources of success and failure, and to look more closely at the ways in which our fascination with the celebrated cases limits our ability to see the larger world in which safety is created and nurtured.

References

Cook RI, Woods DD. Implications of automation surprises in aviation for the future of total intravenous anesthesia (TIVA). *Journal of Clinical Anesthesia*. 1996; 8: 29s-37s.

Cook RI, Woods DD, Howie MB. Unintentional delivery of vasoactive drugs with an electromechanical infusion device. *Journal of Cardiothoracic and Vascular Anesthesia*. 1992; 6: 238-244.

Cooper JB, Newbower RS, Long CD, McPeek B. Preventable anesthesia mishaps: A study of human factors. *Anesthesiology*. 1978; 49: 399-406.

Dominguez C. (1998). Expertise in laparoscopic surgery: anticipation and affordances. In: *Proceedings of Naturalistic Decision Making 4*, Warrenton VA, May 1998.

Dominguez C, Flach J, Lake P, McKellar D, Dunn M. (in press). The conversion decision in laparoscopic surgery: Knowing your limits and limiting your risks. In Shanteau J, Smith K, & Johnson P, (eds). *Psychological Explorations of Competent Decision Making*. New York: Cambridge University Press.

Feltovich PJ, Spiro RJ, Coulson R. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In: Evans D, Patel V, (eds). *Cognitive Science In Medicine: Biomedical Modeling*. Cambridge, MA: MIT Press.

Feltovich PJ, Ford KM, Hoffman RR, (eds). (1997). *Expertise in Context*. Cambridge MA: MIT Press.

Guerlain S, Smith PJ, Obradovich JH, et al. Dealing with brittleness in the design of expert systems for immunohematology. *Immunohematology*. 1996; 12(3): 101-107.

Hollnagel E. (1993). *Human Reliability Analysis: Context and Control*. London: Academic Press.

Holzman RS, Cooper JB, Gaba DM, Philip JH, Small S, Feinstein D. 1995; Anesthesia crisis resource management: Real-life simulation training in operating room crises. *Journal of Clinical Anesthesia*. 1996; 7: 675-687.

Howard SK, Gaba DM, Fish KJ, Yang GS, Sarnquist FH. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine*. 1992; 63: 763-770.

Klein GA. (1998). *Sources of Power: How People Make Decisions.* MIT Press, Cambridge MA.

Leape LL. Error in Medicine. *JAMA.* 1994; 272: 1851-1857.

Moll van Charante E, Cook RI, Woods DD, et al. Human-computer interaction in context: physician interaction with automated intravenous controllers in the heart room. In Stassen HG (ed), *Analysis, Design and Evaluation of Man-Machine Systems 1992*, Pergamon Press. 1993: 263-274.

Obradovich JH, Smith PJ, Guerlain S, et al. Empirical evaluation of the transfusion medicine tutor. *Immunohematology*. 1996; 12(4): 169-174.

Rasmussen J. The role of error in organizing behavior. *Ergonomics*. 1990; 33: 1185-1199.

Reason, J. (1990). *Human Error*. Cambridge, England: Cambridge University Press.

Reason J. (1997). *Organizational Accidents: The Management of Human and Organizational Factors in Hazardous Technologies*. Cambridge, England: Cambridge University Press.

Stewart L, Way LW. Bile duct injuries during laparoscopic cholecystectomy: Factors that influence the results of treatment. *Arch Surg*. 1995; 130: 1123-1128.

Way LW. (1992). Bile duct injury during laparoscopic cholecystectomy. *Ann Surg.* 1992; 215(3): 195.

Weick KE, Roberts KH. Collective mind and organizational reliability: The case of flight operations on an aircraft carrier deck. *Administration Science Quarterly*. 1993; 38: 357-381.

Woods DD, Sarter N. (in press). Learning from automation surprises and going sour accidents. In Sarter N, and Amalberti R. (eds). *Cognitive Engineering in the Aviation Domain*, Erlbaum, Hillsdale NJ.

Woods DD, Johannesen L, Cook RI, Sarter N. (1994). *Behind Human Error: Cognitive Systems, Computers and Hindsight*. Crew Systems Ergonomic Information and Analysis Center, WPAFB, Dayton OH.

Yue L, Woods DD, Cook, RI. (1992). *Reducing the Potential for Error Through Device Design: Infusion Controllers in Cardiac Surgery*. Cognitive Systems Engineering Laboratory Report TR-01-92, The Ohio State University, Columbus OH, January 1992.

Appendix A

List of Participants

Name	*Affiliation*	*Location*
Bagian, James	Environmental Protection Agency	Ann Arbor, MI
Billings, Charles	Ohio State University (formerly NASA)	Columbus, OH
Blim, Jill	National Patient Safety Foundation	Chicago, IL
Boaz, Lev	Israel Health Ministry	Jerusalem, ISRAEL
Bogner, Sue	Institute for Study of Medical Error	Washington, DC
Charles, Sara	University of Illinois at Chicago	Chicago, IL
Cook, Richard	University of Chicago	Chicago, IL
Cooper, Jeffrey	Partners Health System	Cambridge, MA
Diamond, Louis	The MEDSTAT Group	Washington, DC
Dominguez, Cynthia	Wright Patterson AFB	Dayton, OH
Donchin, Yoel	Hadassah Hospital	Jerusalem, ISRAEL
Fischhoff, Baruch	Carnegie Mellon	Pittsburgh, PA
Flach, John	Wright State University	Dayton, OH
Gaba, David	Stanford University	Palo Alto, CA
Goldman, Larry	American Medical Association	Chicago, IL
Goldman, Ronald	Dept. of Veterans Affairs	Washington, DC
Hancock, Peter	University of Minnesota	Minneapolis, MN
Hendee, William	Medical College of Wisconsin	Milwaukee, WI
Hirschhorn, Larry	Wharton Center for Applied Research	Philadelphia, PA
Hollnagel, Erik	Halden Reactor Project	Halden, NORWAY
Klein, Gary	Klein Associates	Yellow Springs, OH
Knox, Eric	MMI Insurance Companies	Minneapolis, MN
Leape, Lucian	Harvard School of Public Health	Cambridge, MA
Margolis, Howard	University of Chicago	Chicago, IL
Mecher, Carter	Dept. of Veterans Affairs	Atlanta, GA
Miller, Charlotte	National Patient Safety Foundation	Chicago, IL
Nance, John	ABC News Aviation Analyst	Seattle, WA
Nozick, Linda	Cornell University	Ithaca, NY
O'Connor, Michael	University of Chicago	Chicago, IL
Palmer, Larry	Cornell Law School	Ithaca, NY
Palmisano, Don	AMA Board of Trustees	Metairie, LA
Parsons, Don	Kaiser Permanente	Washington, DC
Patton, Larry	Agency for Health Care Policy & Research	Washington, DC
Ragan, Michael	CNA HealthPro	Chicago, IL
Ray, Wendy	Kaiser Permanente	Portland, OR
Roberts, Karlene	UC Berkeley	Berkeley, CA

Selby, Joe	Kaiser Permanente	Oakland, CA
Senders, John	University of Toronto	Toronto, Ontario, CANADA
Smith, Phil	Ohio State University	Columbus, OH
Solomon, Ronni	ECRI	Philadelphia, PA
Tasca, Leo	Ministry of Transportation	Downsview, Ontario, CANADA
Vogt, Eleanor	National Patient Safety Foundation	Chicago, IL
Weingart, Saul	Beth Israel Deaconness Hospital	Boston, MA
Woods, David	Ohio State University	Columbus, OH
Xiao, Yan	University of Maryland	College Park, MD
Zipperer, Lorri	National Patient Safety Foundation	Chicago, IL

Appendix B

This appendix contains the talk by Charles Billings, MD, Chief Scientist (retired), NASA Ames, on the lessons learned from incident reporting in aviation. Dr. Billings designed, started and managed the Aviation Safety Reporting System 22 years ago when at NASA's Ames Research Center. His talk framed the discussion on the second day of the workshop. The lessons he abstracted from the aviation experiences represent the best guidance available on incident reporting. Medicine is quite different from aviation in many ways. What proved successful in aviation is not likely to transfer directly and literally to medicine. However, the lessons Dr. Billings has derived are generic and can serve as a guide to develop successful systems in medicine.

Incident Reporting Systems in Medicine and Experience With the Aviation Safety Reporting System

Charles Billings, MD
Institute for Ergonomics
Ohio State University

This is only a brief digest of what I think are some of the most compelling and important issues regarding formalized incident reporting. I hesitate to use the term *systems* to describe the many different approaches to incident reporting. To call them "systems" would dignify them unjustifiably, at least at this point. But I acknowledge that there are various requirements that, in the past, have shaped the systems during design and implementation. The experience with the aviation safety reporting system, in which I developed the ideas that form the basis for this presentation, exemplifies these requirements. The first and most critical requirement for a successful incident reporting system is a demonstrated, tangible, widely agreed upon need for more and better information. If a substantial portion of a community believes that it already knows what needs to be known about incidents, then it is unlikely to give more than lip service to finding out what incident reporting can discover. Strong, widely held consensus that more and better information is needed, is essential for the development of successful incident reporting. The second requirement is for a respected body, one independent of the influences of other stakeholders, to conduct the collection and analysis of data. This is an absolute requirement. Some disagree with this, but many incident reporting schemes have come to grief over the years by being installed in that body which was charged with oversight or in some other body that was subservient to the body charged with oversight, of the community whose activities were being assessed by the incident reporting system.

Two other factors have led to incident reporting schemes coming to grief in the recent past. One is adequate funding to permit expertise to be brought to bear. These systems cannot be run with a couple of clerks and a keypunch operator. For any useful degree of understanding of the reports, incident reporting *requires* expertise at the South end equal to that which was on the North end, that is,

there must be expertise used in evaluating the reports as they are obtained. The other factor necessary is *time*. There must be adequate time to establish the system, to gather data and, more importantly to sell the system, its input, data, information, reports, and conclusions, to those authorities who represent the ultimate recipients of its products.

To some, these prerequisite requirements may seem obvious, but the simple fact is that incident reporting schemes have sometimes failed for want of them.

Let me address very briefly the question of whether incident reporting should be mandated. Those of you who have read the two volumes of materials that were sent in preparation for the conference (and who were still reading carefully by the time you got to the end of them!), will have recognized that the New York hospital incident reporting system is a mandatory system. But the other documents in these volumes you find in several points the observation that incidents are underreported. Well, that will come as a less than an overwhelming surprise to most. Some claim that mandatory reporting is necessary. They say that if you *don't* have mandatory reporting everyone will protect themselves. I add only that if you *do* have mandatory reporting, everyone else will protect themselves. Despite mandatory reports in some locales, the number of incidents reported is small. For example, according to the materials, Carolina had 15 reports in its first year and Colorado had 17 reports in 2 years. I leave it to you to decide whether you think that is adequate or not. I will here make the claim that, in some form, in one way or another, all incident reporting becomes voluntary. It either becomes voluntary because of inertia on the part of reporters, or it becomes voluntary because of constraints within the establishment and the environment, or it becomes voluntary because hospitals (and there are at least one or two in your two source books) decide that they are not required to report this particular event because of the fine print in that particular incident reporting regulation or statute.

Underreporting is a recognized problem. But I'm not at all sure that that is the *critical* problem. Staying within the medical context and referring to the materials from your source books, there are enough reports of mishaps with potassium chloride, lidocaine, vincristine and other drugs and devices to have made it very clear that a problem with these exists. The information that these events occur is already present. We may well ask what it is that keeps us from making progress on safety, given that we already know about the existence of these problems. What is added by more formal, elaborate (and expensive) incident reporting? We may hypothesize that if the events had become known to a central organization, there would have been discovery of trends and systematic evaluation of what was going on that, in turn, would have led to some unspecified but important change. But it is clear, even at this stage, that the hypothetical, safe, non-punitive incident reporting system is not the primary obstacle to making progress on safety in medicine. Existing agencies can take action once a problem is adequately defined, explored, and explained (and, I would have to add, publicized). A central question facing us is not really how many there *are*, but *how many is enough*. That is to say, there are already many signals that point to a variety of failures. Part of the consensus that needs to be formed for successful incident reporting is consensus about what is a sufficiently strong signal to warrant action. The sourcebook materials suggest that such a consensus remains to be developed.

Stakeholders influence the likelihood of developing consensus. Now once again I've dipped into the sourcebook to the CEO's retrospective report on the Ben Kolb case, which is volume 1 tab 4. I've added two here because they were spoken of elsewhere but I found that the CEO listed a rather substantial number of stakeholders in that two-page report.

The CEO of the organization	Risk management department
The governing body	The insurer(s)
The staff of the organization	The media
The staff directly involved	The medical examiners
Physicians and nurses	Outside consultants
Support personnel	Accrediting organizations
The patient	The union(s)
The family	Advocacy groups
"The government(s)"	The attorneys

That, if you will forgive me, is a lot of people, organizations, interests, stakes and conflicting goals. If I am prepared to argue to you that incident reporting requires consensus, it seems reasonable to ask whether consensus can be reached among these stakeholders on *anything* including whether it is day or night, whether the sun is moving to the west or the east.

It will surely be asked by many why consensus among these stakeholders is necessary. What does it buy if you have it? First, it buys a substantial number of people and organizations as advocates for incident reporting and the system that supports it. This is essential to keeping the system a working, living entity in a contentious, politically driven environment. Second, it buys the participation of those people who can, if they wish, do something with the information you produce.

It might be useful to turn the question of consensus on its head. What does it cost you if you don't have it? At this point I will diverge just slightly to answer Richard Cook's question, posed at the beginning of this conference. He asked what critical failures we now recognize after a lifetime of research. This led me to think back on the history of the aviation safety reporting system and to identify the most significant failure in its development. The NASA Aviation Safety Reporting System was established in response to the cries of virtually everybody in the community, directly at the order of the FAA Administrator working through the Administrator of the National Aeronautics and Space Administration. NASA was chosen because we represented a respected and presumably objective third party. We were given one month to establish the system, guide it through about as many stakeholders as are shown in the table, and present it to the Administrator as a done deal. We (or, rather, I), in my naiveté made an assumption. I assumed that since the chief of the FAA was asking for it, the FAA wanted it. This was a bad mistake.

Over the past 21 years, the NASA Aviation Safety Reporting System has had the support of virtually everybody in the United States aviation community to a greater or lesser extent with one notable exception. The exception is the people under the FAA Administrator—and there are roughly 22,000 of them—whom he

did not consult before he came to NASA and asked for an incident reporting system. We do well to remember that a primary issue is who may be hurt by reporting. This is especially of concern where use of immunity (and, originally in the ASRS, transactional immunity as well) is a prominent feature.

What was the largest segment of the FAA aside from the air traffic control system? Divisions involving regulations and enforcement. Whose enmity did we earn the day this thing was announced? Those who had to make it work in the community. That is the worst mistake I made in 40 years.

So the absence of consensus about the need for and characteristics of incident reporting was a critical flaw in the development of ASRS. What does it cost you if you don't have consensus? It costs you in passivity; resistance to acting on a lot of the recommendations derived from data received; delays in implementing; even ridicule. "You know it isn't a problem," was one form this passivity and resistance took. Another was "You guys just don't understand." Consensus is critical and it must include all the stakeholders, not just a few or a special team or a division or an agency or a company.

I also want to point out that consensus isn't enough. It's necessary but it is not sufficient. Incident reporting also requires *understanding* and that is even tougher to establish. Incident reports are unique sets of data. Each incident is unique and not easily classified or pigeonholed. Generalizations may be possible in retrospect, given enough detailed data and enough understanding of the data. But this means understanding details of the task, the context, the environment, and its constraints. This is why you have got to have experts looking at reports. Simply constructing taxonomies is grossly insufficient and it permits only counting of incidents that fall under phrase a, b or c of the taxonomy.

Counting incidents is a waste of time. Why? Because incident reporting is inherently voluntary. Because the population from which the sample is drawn is unknown and therefore can not be characterized, and because you lose too much information and gain too little in the process of condensing and indexing these reports unless you do what we were fortunate enough to do blindly, and that is keep all the narratives. Every ASRS report is in fact a narrative rather than a categorization. And the ASRS keeps every word, except those necessary to be discarded to de-identify the reports. The evaluation of incidents in such a system requires an understanding of all that. A deep understanding.

Let me give you an example. We got a report about 2 years after the ASRS started which, paraphrased, said "I had a frightening experience this morning. I took my airplane, a Lockheed 1011, off at Los Angeles headed for Vancouver, and it took full right aileron to keep it in the air. The airplane wanted very badly to roll sharply to the left." Fortunately, it was a cold day at Los Angeles, the load in the airplane was light, the pilot, carrying full right aileron all the way and with help from the first officer, managed to struggle around the airport and get it back on the ground at Los Angeles and no one was hurt. Whereupon it was found by the flight engineer that the two outboard spoilers on the right wing were both fully extended and had been since the airplane left the ramp.

Why was this not caught? Neither of those two spoilers carries sensors for spoiler deflection, so the crew, once the airplane was powered up, had no indication in the cockpit that those devices were extended. You can't see them from the cockpit. The airplane was powered down at the time the first officer made the walk around, and in an all-hydraulic airplane, controls can be in any position whatever when there is no hydraulic power on. We had three other similar reports after that, but we really didn't wait for a second report. We immediately got why that happened and reported it, de-identified, to those people we thought needed to know about it. As a matter of fact, the reporting back was conducted by telephone and that incident was handled initially within 24 hours.

Counts are not what you are after. You can't, in a voluntary system, determine the incidence or prevalence of a problem. But if you get two or three reports like that, you can be pretty sure that you have a problem, and that is what you are after in this kind of a system. On the other hand, having extolled the kinds of things incident reporting can do for you more briefly than I usually do, incident reporting is not the whole answer. I think this is particularly germane to your considerations here. In aviation accident reporting is mandatory, and accidents are very precisely defined. Yesterday we discussed the problems associated with defining and characterizing what an accident is. In aviation we have a much easier job, and the distinction between an accident, which must be reported, and an incident is easier to make than is the case in health care. Keep in mind that incident reporting in aviation is voluntary under most circumstances. I*ncident reporting and accident reporting are not substitutes one for the other. They are complementary.*

Incident reports, properly interpreted, provide new knowledge. It is important to remember that this is all they provide. They are not a panacea. They only provide knowledge about what is going on in a particular domain or area of purview. The incident reports themselves tell you little. They generally do not tell you how it could be done better. The analysts, looking at many incidents, may or may not be able to tell you how it may be done better. They are more likely to know, in many cases, more than the incident reporters. Certainly they are more likely, over time, through reading these things, to gain some experience and understanding. But like descriptive epidemiology anywhere else, incident reports can provide only descriptions of phenomena. Analytic studies and other research remain critical to a full understanding of the phenomena.

Acting on the new knowledge that comes from combining incident reporting with analytical studies and detailed research is not and should not be within the purview of an incident reporting system. I emphasize this because too many people have thought that incident reporting was the core and primary component of what was needed. These people thought that simply from the act of collecting incidents, solutions and fixes would be generated *sui generis* and that this would make safety better. Although much is unclear about incident reporting systems, this one fact is quite clear: incident reporting is only one component of what you need. Using new knowledge gained from these systems must be the responsibility of the stake-holders I have listed. This says that the use of this knowledge to spur new analysis, new research, to guide regulation, to inform management decision making, to change performance, must be the responsibility of everybody else in the domain. It must not be the job of the people who run the incident reporting

system, who have got to remain objective and disinterested in order to do this job as it needs to be done.

The incident reporting system must remain an objective and disinterested third party to retain its effectiveness. Otherwise, sooner or later, it will be accused, properly, of bias. There are enough biases we cannot get rid of; it clearly does not make sense to add more. So, I ask you to consider among other things, what should be reported. Is harm required? How much harm? What kind of harm? Is an adverse event—and that is the term Kaplan and his colleagues are using in Texas—determined by processes or by outcomes? Can incidents be differentiated from accidents in these settings? The difference is very critical with respect to legal liability. Unfortunately as we see the increasing trend, referred to in your sourcebook, toward criminal prosecution as well as civil liability, this question becomes increasingly important. It becomes increasingly important because it is entirely possible that if someone decides to go after a criminal indictment following such an incident or accident or mishap, you may be harboring evidence. That is not a particularly desirable position for the hapless expert working in one of these outfits to be in. I believe the legal term for it is obstruction of justice. This was a serious problem in Canada, where a violation of what in the US is called an air navigation order is not only a civil violation, it is also a criminal offense. The people who work within the Canadian confidential air safety reporting system had to be specifically immunized by Parliment to allow them to do this task. That would be true in any nation that was governed by the Napoleonic code. It is increasingly a potential threat in the common-law nations as well, given what I said yesterday about corporate manslaughter, corporate crimes involving damage or injury to people. So does it matter? Twenty years ago I would certainly have said, "Probably not." Now I would say, "Yes, I think it probably does." I am not sure yet of the way to get around this, and I'm doing a fairly extensive study at the present time to try to figure out how much it matters. But there is certainly, in aviation at least, an increasing trend toward imprisoning or invoking other criminal sanctions against people for offenses committed in the course of their employment in this industry.

Should records ever be protected? There are certainly going to be differences of opinion with respect to that. Can they ever be protected? One of the most treasured bits of data within the aviation industry, are the cockpit voice recorder (CVR) tapes. The police in New Zealand, after an accident there in 1995, subpoenaed, then sought a court order, then went to trial to get the CVR tape from that accident in connection with prosecution of the pilot and co-pilot, one of whom lived and the other of whom did not. The police prevailed at the appellate level and were given the CVR tape as criminal evidence. That has led to a very substantial effort in New Zealand to decide what are the limits of evidence gathering in this environment. And, as a matter of fact, the New Zealand Law Commission is looking into that at the present moment. This is a very serious question. Another question, related to it, is should the analytical records of an incident reporting system be protected? Sufficiently detailed, these analytical records may make it almost unnecessary to get to the raw data. In this country, thus far, raw data, from the aviation safety reporting system once de-identified has been considered hearsay and therefore inadmissible in the vast majority of cases.

That may be the balm that can be spread upon the wounds in some of these other nations within the common-law states but in any event, this question remains an important one.

And finally, is an incident reporting system possible in this, that is to say, the medical environment? Thinking of the number of stakeholders involved, thinking of the enormity of the hypothesized problem, I think that we all need to reserve judgment. We need to have more evidence, gathered by different means—converging evidence to bring to bear on this question. Until this problem is scoped, neither the National Patient Safety Foundation nor anyone else is going to be able to figure out how to tackle it, so I think that that should clearly be one of your front-end research objectives. But however large the problem, the diversity amongst the stakeholders is very worrying. Now in the supplementary sourcebook, I've included some view graphs describing the NASA Aviation Safety Reporting System. A couple of those deal with our advisory committee, which represents a broad segment of the large, complex, diverse, aviation industry. But the size, complexity, and diversity of aviation are simply not on the scale of medicine; in fact, aviation is nowhere near. I would offer you some hope that it may be possible to define areas within medicine in which there is a somewhat smaller constituency, a somewhat smaller group of stakeholders and within which, therefore, the problem may be slightly more tractable than it will be if you decide that your purview is all of medicine. I am not sure I would know at this point, even after 20-plus years of experience with this business, how to design a system for medicine. I think I could perhaps design a system for some subsegments of medicine, that are more tightly circumscribed, but I hasten to add that I am not even sure of that.

These are some of the questions you are going to have to face as you begin to think about incident reporting systems in this particular domain. They make success in medicine much less certain, in contrast with the undoubted success that incident reporting systems have had in the aviation environment and a very few others. There are many other detailed issues to consider in developing incident reporting within medicine, issues that you will need to address in detail, should you chose to pursue this line of work. All these other questions deserve attention. But you will find, I think, that the ones I have presented here today are among the largest, most important ones that will be critical to the success or failure of any new incident reporting system in medicine. I think those kinds of questions need to be thought about very carefully, deliberately, intelligently, and thoroughly before beginning.

***Question 1: Would you comment about your efforts over the years to keep the ASRS alive*?**

Well, that was the first thing we had to do. In some ways, this was actually fairly straightforward because the industry, by and large, recognized that they had a problem. The Administrator of the FAA had a political problem, but this is not particularly germane in our discussion today. The larger aviation industry, including all the stakeholders, had a real problem and recognized it. So this particular seed fell upon reasonably fertile ground. We got along reasonably well initially until we began to recognize the depths of the disinterest of the FAA working staff in this program. We did everything we could to interest them. We went out and we

talked to as many of them as we could get to. We worked within FAA headquarters to try to get them to understand what they had put us into.

This was not entirely successful, but we got along by and by fairly well, until political problems arose concerning a freedom of information act request asking for an evaluation and discussion of air traffic congestion in terminal areas. The ASRS was obligated to respond to this, and it did. We put together a report whose principal finding of concern was that near mid-air collisions appeared from our data to be most likely in the highest density terminal control areas in the United States. The next most likely areas for near mid-air collisions were what were called Airport Radar Service Areas. These were the terminal areas around secondary but still busy airports. Near mid-air collisions were less likely to occur anywhere else.

This would probably not surprise you all much and it did not surprise us. But, it was of course a direct affront to the agency which regulates air space in the United States, and which had been saying, as part of its campaign to establish some 10 or 12 new terminal control areas, that they were the safest place in the air. The internal political response to this report was complex but amounted to an effort to remove the legal immunity that protected reports to the ASRS. Although this was forestalled by the intervention of Congress, it is clear in retrospect that the survival of the system as a working entity depended on the consensus within the community that had been reached early in the development of the system. It is one example of the importance of stakeholders who are directly and actively involved.

Question 2: Does the fact that pilots are at personal risk when flying have something to do with the success of the system? Is that a reason that you think this system might work much better in aviation than it might be made to work in medicine?

Well, it is certainly true that pilots are normally the first people at the scene of an accident. But the answer in my view is no, it is not the sense of personal risk that has driven pilots to participate in the system. I read the first 30,000 of these reports that came in. I have not read the ensuing 340,000 or so. One is continually amazed by the number of people who do far more than they need in order to qualify for immunity under this system. People write pages of descriptions, they send tapes, they volunteer to come to our offices. They make clear in a number of ways that they want us to understand, in all its rich detail, the complexity of the incident in which they had been involved and which they had already reported by mail. One indication of the sincerity and dedication of these people is that some have had such close interaction that they have become personal friends. There is no question about the motivation of the pilot community in general with respect to safety issues, none whatever. The reports are not grudging acknowledgement or pro forma filings but rather quite rich and human descriptions of troubling, often frightening events. I believe that the reporting to this system is motivated not by the sense of personal risk that attaches to flying but rather from two major factors: (1) the sincere interest in improving safety by identifying hazards and (2) the sincere (and, in my view, well grounded) belief that the system to which they are reporting uses that information productively and deliberately to improve safety rather than simply as a means of counting failures.

Parenthetically, I know that the system has been "gamed" in the course of adjudication of labor issues. I think it would be very surprising if it were not. But that becomes relatively easy to pick out; one notices it. One occasionally gets some very good information out of those reports incidentally, because some of that gaming is over issues that are real and substantive. The fatigue and jet lag issue is one that I think of immediately although that is by no means the only one. But once again you have to have a human—a smart human and a motivated human—who understands and has been there at the sharp end of the process from which reports come as well as at the analytical end. And I do not believe there is any substitute for that.

Question 3: Could you comment on expertise issues in the procedure for examining incidents and looking for patterns?

Pilots get rusty after about so long; as you know, many of our analytical people are retired pilots. We have instituted a set of rotations for these people, whereby they would come and work for us for a period of time and then they would move on and we would replace them with others coming out of the cockpit. Now that has been in place for probably fifteen years now, and it is also true with respect to our air traffic controllers. No one can stay adequately familiar with the niceties in an industry this complex unless he or she has been actively working in it until comparatively recently.

Interestingly, although we have been overwhelmed at times by the number of reports, the system for handling those reports has changed remarkably little. They are all hand-read by an expert who is appropriate to the field of consideration in the report. The reports are still identified at that stage and are being handled as classified material, so that if more data is required we can get it directly from the reporter. We still have the pilot's name, phone numbers, things of that sort. We will attempt to call the pilot back and discuss in more detail the incident that was reported and sometimes secondary issues as well, before the reports are de-identified. We do classification on a number of classification fields. The reports are processed, the narratives are de-identified with respect to person, flight numbers, and things of that sort, and then those narrative reports are keyed directly into a data base as well as of 60 or 70 fields of coded data. So the classification system, which is primarily used for indexing, provides access to the narrative. It is also possible to search the narrative fields by words, by phrases, and considerable amount of research has been done using that, as well. We would have lost all of that had we destroyed the narrative.

Question 4: What does it cost to run this system now?

To the best of my knowledge it's about two million dollars per year for about 35,000 to 40,000 reports.

Question 5: Currently in the medical field, legal requirements for discoverability drive access to these incident reports. In Louisiana and Arizona, for instance, between 50% and 100% of incident reports are discoverable in civil cases. This has created a substantial incentive to make these reports narrow and short. There is a strong emphasis, stated explicitly in most cases as directions to the reporting individual, to limit what is

described to objective rather than subjective information. For example, regarding a fall in the hospital, what people are instructed to record and report is the bare fact that someone slipped and fell. If the individual said, "I slipped on water" the incident reporter can put that in quotes but needs to avoid concluding that they slipped on water. Thus these reports are limited to observations of the scene, noting the times, locations and so on. Are the ASRS reports specifically limited in those ways—that is, narrow, succinct, objective and without any opinions about the incident or how it occurred? Will such a system produce results similar to those obtained via the ASRS?

Absolutely not. We rely on preservation of this material as classified material initially, on rapid de-identification, and on blurring enough of the details of such a report so that it would at least be difficult to introduce it as evidence. But we also need the reports of people, of pilots, who we understand are experts in their field and whose opinions, thoughts, and observations we value. We do not try to limit them or channel their descriptions in a particular way. Indeed, it would surely limit the value of the reports if we were to artificially constrain the narratives to eliminate any conclusions or opinions from the pilots. But these reports are data, and we all recognize that data has biases and limits and the important thing is to try to understand these. This goes back to the problem of throwing away data when you throw away the narratives—counting doesn't work, that is not what you are after in incident reporting. Limiting the narratives in the way you describe would be throwing away the data even before you got it, a serious mistake in my view.

Now with respect to de-identifying the events in the database—that can be a bit hard. I think particularly of this city in which we find ourselves, and arriving here early one morning shortly before I retired from NASA, to see a 747 sitting ingloriously in the dirt, between Runway 10-Right and the taxiway adjacent to it. Now that was not, as it happened, an accident. The monetary value of it was not high enough to require defining it that way, partly because it was soft mud and there was little damage to the plane. But I would have hated to try to de-identify that report sufficiently so that it would be in the database but unidentifiable. A 747 is one big moose when it is sitting out there where everyone can see it.

Appendix C

Sourcebook: Workshop on Assembling the Scientific Basis for Progress on Patient Safety. Volume I

Section A. Celebrated Cases

The "celebrated" cases are medical accidents that have attracted a great deal of attention in the press and from various stakeholders. A variety of reactions from healthcare professionals, regulators, and the public appear in these articles. Most of the material included on these cases comes from newspaper articles. Also included is a more discursive document on the 'wrong leg' case that indicates the perspective of the state regulatory body and the rationale for the revocation of the surgeon's license. Also included are some cases demonstrating the recent trend towards treating medical accidents as criminal events (usually manslaughter).

A-1 (Willie King wrong leg case; Dr. Rolando Sanchez)

Agency for Health Care Administration vs. Rolando Roberto Sanchez, M.D. Fla. Div. of Admin. Hearings. Recommended order. Case no. 95-3925

Associated Press.
Tampa hospital had earlier surgical slipup. *The Ledger* (Lakeland, Fla). March 8, 1995; news:5B.

Wrong-leg surgeon cut: toe for two. *The Commercial Appeal* (Memphis). July 19, 1995; news:7A.

Bartlett R. Hospital safety law proposed: bill would compile data on deaths, injuries, infections. *The Tampa Tribune*. February 9, 1996; Florida/metro:6.

Clary M. String of errors put Florida hospital on the critical list. *Los Angeles Times*. April 14, 1995; part A:1.

Hagigh J. Family sues surgeon, UCH. *St. Petersburg Times*. January 26, 1996; Tampa today:4B.

Landry S.
Expert: King error a group effort. *St. Petersburg Times*. September 14, 1995; Tampa today:1B.
Fatal error earns suspension. *St. Petersburg Times*. February 3, 1996; Tampa today:3B.
Surgeon fined, suspended 6 months for error. *St. Petersburg Times*. December 3, 1995; national:1A.

Leisner P. Surgeon says it was too late to stop amputation on wrong leg. Associated Press, September 14, 1995.

Mahan M. This time, hospital's error is fatal. *St. Petersburg Times*. March 12, 1995; Tampa Bay and state:1B.

Oppel S. Medical board examines fatal error. *St. Petersburg Times*. October 26, 1995; Tampa Bay and state:1B.

Palosky CS.
Doctor's penalty reduced. *The Tampa Tribune*. December 3, 1995; nation/world:1.
Surgeon fights penalty. *The Tampa Tribune*. November 8, 1995; Florida/metro:6.
Wrong-amputation surgeon makes return. *The Tampa Tribune*. February 29, 1996; Florida/metro:6.

Ripley J. Amputated foot worth more than $1-million. *St. Petersburg Times*. May 12, 1995; Tampa Bay and state:1B.

Rosen M. Report: hospital cut safeguards. *St. Petersburg Times*. April 7, 1995; national:1A.

Ryan P. State slaps surgeon: agency says doctor "serious danger" following 2 amputation complaints. *The Tampa Tribune*. July 15, 1995; nation/world:1.

Shaver K, Carlton S. Right foot amputated. *St. Petersburg Times*. March 8, 1995; around town:1B.

Stanley D.
- Amputee recovering after wrong leg taken. *The Tampa Tribune*. February 28, 1995; nation/world:1.
- Bill seeks hospital disclosures: sponsors want data on injuries inflicted upon patients to be made public. *The Tampa Tribune*. March 4, 1995; Florida/metro:1.
- Federal agency tells hospital to fix problems. *The Tampa Tribune*. March 30, 1995; Florida/metro:1.
- Hospital cites new safeguards. *The Tampa Tribune*. March 2, 1995; Florida/metro:1.
- License won't assure surgeon job: Rolando Sanchez may be allowed to practice again, but he must convince hospitals and insurers to take him back. *The Tampa Tribune*. January 17, 1996; nation/world:1.
- Mistake found during surgery: testimony begins in appeals hearing for doctor who amputated wrong foot. *The Tampa Tribune*. September 13, 1995; nation/world:1.
- Surgeon faced previous claim: the same physician recently was involved in a wrong-foot amputation. *The Tampa Tribune*. March 2, 1995; Florida/metro:1.
- "That's the wrong leg:" victim tries to cope with botched surgery. *The Tampa Tribune*. March 10, 1995; nation/world:1.
- UCH reduces number of surgeries. *The Tampa Tribune*. April 7, 1995; nation/world:10.

Surgeon suspended for cutting off wrong leg. *The Herald* (Glasgow). December 4, 1995:6.

We don't need this kind of doctor [editorial]. *The Tampa Tribune*. July 20, 1995.

Wilson M. Surgical horrors. *Chicago Tribune*. March 10, 1995; news:7.

A-2 (Betsy Lehman)

Altman LK.
- Committees find signs of weak leadership at Dana-Farber. *The New York Times*. October 31, 1995; sect C:5.
- Hospital is disciplined by agency after errors. *The New York Times*. April 16, 1995; sect 1:12.

Associated Press. Settlement reached in overdose lawsuit. *The New York Times*. August 25, 1995; sect A:20.

Estrich S. Why protect doctors? *USA Today*. March 30, 1995; news, counterpoints:13A.

Goodman E. The diagnosis: losing patients, losing faith. *The Plain Dealer*. April 2, 1995; perspective:3C.

Knox RA.
- Chemotherapy deaths spur safer methods. *The Boston Globe*. December 23, 1996; metro/region:B1.
- Dana-Farber, doctor face malpractice suit. *The Boston Globe*. April 6, 1995; metro/region:32.
- Dana-Farber head quits 2d post, vows changes. *The Boston Globe*. May 26, 1995; metro/region:1.
- Dana-Farber probe widens: three suspended from patient care. *The Boston Globe*. April 1, 1995; metro/region:13.
- Dana-Farber puts focus on mistakes in overdoses. *The Boston Globe*. October 31, 1995; metro/region:1.
- Dana-Farber studies pattern in overdoses. *The Boston Globe*. March 26, 1995; metro/region:1.
- Dana-Farber tests signaled an overdose, records show. *The Boston Globe*. May 2, 1995; metro/region:1.

Dana-Farber wins near-perfect score: changes follow 2 patient overdoses. *The Boston Globe*. June 5, 1996; metro/region:44.
Doctor's orders killed cancer patient: Dana-Farber admits drug overdose caused death of Globe columnist, damage to second woman. *The Boston Globe*. March 23, 1995; metro/region:1.
Hospital's record of sympathy faulted: says response to distress factored in death. *The Boston Globe*. July 16, 1995; metro/region:21.
Licensing board reprimands 3 Dana-Farber pharmacists. *The Boston Globe*. September 26, 1996; metro/region:B2.
Media spotlight helped spur change, shook up patients, staff. *The Boston Globe*. December 26, 1995; metro/region:20.
New Dana-Farber head vows "never again." *The Boston Globe*. September 16, 1995; metro/region:1.
New policies put in place. *The Boston Globe*. December 26, 1995; metro/region:20.
Overdoses still weigh heavy at Dana-Farber: more than year after tragedy, cancer institute works to balance research mission, crucial details of patient care. *The Boston Globe*. December 26, 1995; metro/region:1.
President of troubled Dana-Farber steps down. *The Boston Globe*. September 13, 1995; metro/region:1.
Response is slow to deadly mixups: too little done to avert cancer drug errors. *The Boston Globe*. June 26, 1995; science & technology:29.
State cites Dana-Farber failures: quality assurance program faulted. *The Boston Globe*. May 31, 1995; metro/region:1.
State probe cites lax management at Dana-Farber. *The Boston Globe*. May 25, 1995; metro/region:1.
State regulators vow multi-agency probe of Dana-Farber overdose. *The Boston Globe*. March 29, 1995; metro/region:21
Survivor's spirit beats a chemotherapy error. *The Boston Globe*. December 17, 1995; metro/region:1.
Top Dana-Farber doctor steps down. *The Boston Globe*. May 11, 1995; metro/region:1.

Knox RA, Blanton K. Overdoses cloud Dana-Farber bonds. *The Boston Globe*. March 24, 1995; economy:77.

Knox RA, Golden D.
Dana-Farber turmoil seen. *The Boston Globe*. May 28, 1995; metro/region:1.
Drug dosage was questioned: Dana-Farber pharmacist sent order back to doctor in breast cancer case. *The Boston Globe*. June 19, 1995; metro/region:1.

Kong D.
Dana-Farber given the OK for Medicare after correction. *The Boston Globe*. August 3, 1995; metro/region:32
State faults Dana-Farber: says center failed to tell promptly of two overdoses of cancer drug. *The Boston Globe*. March 24, 1995; metro/region:1.

Lasalandra M. Dana-Farber workers probed in death. *The Boston Herald*. October 31, 1995; news:12.

Lessons of a life lost [editorial]. *The Boston Globe*. March 24, 1995; editorial page:18.

Overdose survivor forgives cancer institute in Boston. *Chicago Tribune*. December 18, 1995; news:1.

Pertman A. Betsy A. Lehman, 39: Globe columnist on personal health [obituary]. *The Boston Globe*. December 4, 1994; obituaruies:56.

Spingarn ND. Doctors can prevent errors just by listening [letter]. *The New York Times*. July 25, 1995; sect A:14.

Stein C. Doctor sues Farber Globe in overdose case. *The Boston Globe*. February 3, 1996; metro/region:1.

Trafford A. The empathy gap. *The Washington Post*, August 29, 1995; health:Z6.

Warsh D. Molecular medicine vs. bedside manner in the Lehman case: economic principals. *The Boston Globe*. September 17, 1995; economy:79.

A-3 (Vincent Gargano)

Associated Press. Cancer patient dies after chemotherapy overdose. June 15, 1995.

Fatal chemotherapy overdose suit filed. *Chicago Tribune*. June 21, 1995; metro west:3.

Fegelman A.
Chemotherapy fears increase after death: dosage checks failed, U. of C. officials admit. *Chicago Tribune*. June 16, 1995; Chicagoland:1.
U. of C. cancer patient dies of chemotherapy overdose. *Chicago Tribune*. June 15, 1995; Chicagoland:2.

The New York Times. Cancer patient in Chicago dies after chemotherapy overdose. *The New York Times*. June 18, 1995; sect 1:17.

Phillips SE. Editorial's outrage misplaced [letter]. *Modern Healthcare*. July 17, 1995; letters:32.

Spate of errors tarnishes prestigious reputations [editorial]. *Modern Healthcare*. June 26, 1995; editorial:42.

A-4 (Ben Kolb)

Cox News Service.
"Medical accident" cited in boy's death. *St. Petersburg Times*. January 12, 1996; Tampa Bay and state:1B.
"Mistake" in surgery led to boy's death, Fla. hospital admits; family settles with "no hard feelings." *The Phoenix Gazette*. January 12, 1996; news:A32.

Examining Errors in Health Care: Developing a Prevention, Education and Research Agenda. A case study: how does the health care system respond to an injury caused by error? Symposium. Rancho Mirage, Calif. October 13-15, 1996.

Graulich H. Hospital admits syringe mix-up killed boy. *The Palm Beach Post*. January 11, 1996; A sect:1A.

Krupnick, Campbell, Malone, Roselli, Buser, Slama & Hancock, A Professional Association. In memory of Ben: October 13, 1988 - December 14, 1995.

MMI Companies, Inc. Developing a crisis management and communication plan. *Risk Modification Strategies*. July, 1995.

Reeder J. St. Lucie boy's death ruled bad reaction to epinephrine. *The Palm Beach Post*. December 16, 1995; local:1B.

Trontz I. Hospital's admission of fatal error cited as model. *The Palm Beach Post*. November 18, 1996; A sect:1A.

A-5 (Libby Zion)

Agus C. Zion nurse didn't want to "hang" doc. *Newsday*. December 2, 1994; news:A33.

Andrews L. "I was juror no. 6, the lone dissenter in the Libby Zion case [letter]. *The New York Times*. February 21, 1995; sect A:18.

Asch DA, Parker RM. Sounding board: the Libby Zion case. *The New England Journal of Medicine*. March 24, 1988; 318:771-775.

Associated Press.
Doctors faulted in 1984 death of 18-year-old. *The New York Times*. May 26, 1990; sect 1:27.
Judge halts regents censure of two doctors in Zion case. *The New York Times*. June 23, 1990; sect 1:27.

Belkin L. Effects are unclear in limits on medical residents' hours. *The New York Times*. February 23, 1993; A:1

Cantor C. Hospital residents pressure for change. *The New York Times*. August 9, 1992; sect 13:1.

Carmody D. New York Hospital disputes allegations on patient's death. *The New York Times*. January 15, 1987; sectB:2.

Cohen T. About health; interns' hours: a red herring. *Newsday*. June 23, 1989; viewpoints; New York forum:80

Colburn D. Medical education: time for reform? after a patient's death, the 36-hour shift gets new scrutiny. *The Washington Post*. March 29, 1988; health:Z6.

Collins D. A father's grief, a father's fight. L*os Angeles Times*. February 1, 1995; life & style, part E:1.

Daley S. Hospital interns' long hours to be reduced in New York. *The New York Times*. June 10, 1988; sect A:1.

Douglas RG Jr. Zion case verdict vindicates training system [letter]. *The New York Times*. February 14, 1995; sect A:18.

Dubin J. Zion case verdict vindicates training system: on the evidence [letter]. *The New York Times*. February 14, 1995; sect A:18.

French HW.
A limit on doctors' hours is debated. *The New York Times*. April 5, 1989; sect B:3.
In overhaul of hospital rules, New York slashes interns' hours. *The New York Times*. July 3, 1989; sect 1:1.

Fugh-Berman A. Exhausted doctors don't serve patients well. *The New York Times*. March 3, 1995; sect A:26.

Handelman S. Student's death haunts U.S. medical establishment. *The Toronto Star*. opinion:A29.

Hershenson R. New rules pose conflict for interns. *The New York Times*. February 14, 1993; sect 13WC:1.

Hoffman J.
Doctors' accounts vary in death of Libby Zion. *The New York Times*. January 1, 1995; sect 1:31.
In the Zion case, a doctor and profession on the stand. *The New York Times*. November 19, 1994; sect 1:27.
Judge sets aside use of cocaine as part of verdict in Zion case. *The New York Times*. May 2, 1995; sect B:1.
Jurors find shared blame in '84 death. *The New York Times*. February 7, 1995; sect B:1.

Japenga A. Endless days and sleepless nights: do long work schedules help or hinder medical residents? *Los Angeles Times*. March 6, 1988; view, part 6:1

Kleinfield NR. Doctor's criticism in '84 case strains hospitals' bonds. *Newsday*. August 5, 1991; sect B:1.

Kocieniewski D. Libby Zion doc admits medical team failure. *The New York Times*. November 17, 1994; news:5.

Landers A. Exhausted resident questions ability to make decisions. *The Houston Chronicle*. November 2, 1997; lifestyle:4.

Libby Zion [obituary]. *The New York Times*. March 6, 1984; sect B:10.

Mitler MM, Cetel MB. Why the doctor may fall asleep on your case. *Los Angeles Times*. April 9, 1989; opinion, part 5:5.

Myers MF. When hospital doctors labor to exhaustion. *The New York Times*. June 12, 1987; sect A:30.

Myerson M. Long hours of medical residents improve care [letter]. *The New York Times*. February 23, 1995; sect A:22.

Okie S. New prescription for medical interns? controversial N.Y. plan would limit hours of hospital shifts. *The Washington Post*. June 30, 1987; sect 1:A1.

Prosecuting doctors [editorial]. *The National Law Journal*. September 16, 1985:12.

Roark AC. Lack of sleep not impairing, study finds. *Los Angeles Times*. September 23, 1988; part 1:3

Ronches J. Must we squander the legacy of Libby Zion? *Newsday*. January 3, 1995; view points:A24.

Rovner S. The battle over limiting residents' hours. *The Washington Post*. October 16, 1990; health:Z9.

Sack K. Appeals court clears doctors who were censured in the Libby Zion case. *The New York Times*. November 1, 1991; sect B:2.

Scott G. Court upholds limits on hours for medical residents. *Newsday*. December 30, 1990; news:18.

Segal MM, Cohen B. Hospitals' junior doctors need senior backups. *The New York Times*. June 8, 1987; sect A:18.

Some limits supported for intern work hours. *The New York Times*. August 21, 1987; sect B:3.

Sullivan R.
Hospital admits fault in patient's death. *The New York Times*. March 24, 1987; sect B:3.
Third doctor is charged in death. *The New York Times*. August 11, 1987; sect B:3.

Two of Zion's doctors accused in her death. *The New York Times*. May 1, 1987; sect B:2.

Weiner E. Panel clears Zion's doctors of wrongdoing in her death. *The New York Times*. September 23, 1989; sect 1:28.

Wicker T.
Blaming the system. *The New York Times*. February 4, 1987; sect A:27.
Doctors in the dock? *The New York Times*. December 23, 1985; sect A:17.

Zion S. Doctors know best? *The New York Times*. May 13, 1989; sect 1:25.

A-6 (Dr. Gerald Einaugler)

Associated Press.
Doctor going to prison for fatal medical mistake. March 16, 1995.
Doctor's mistake lands him in city jail convicted of trying to hide fatal error. *The Record (Bergen Record Corp)*. March 17, 1995; news:A14.
Doctor's order kills patient: physician goes to jail. *The Commercial Appeal (Memphis)*. March 17, 1995; news:6A.

Bayles F. Criminal indictments against doctors seen as "warning shot." April 12, 1995.

Bensel FP, Goldberg BD. Prosecutions and punitives for malpractice rise, slowly. *The National Law Journal*. January 22, 1996; in focus, health care:B7.

Bowles P. Doctor does prison time on weekends: L.I. internist serves sentence in B'klyn nursing-home death. *Newsday*. May 20, 1997; news:A8.

Cohen J, Shapiro R. The criminalization of medical malpractice. *New York Law Journal*. August 28, 1995; outside counsel:1.

Crane M.
Could a clinical mistake land you in jail? the case of Gerald Einaugler. *Medical Economics*. February 21, 1994; 71 no.4:46.
Prosecuting doctors for bad outcomes. *Medical Economics*. August 21, 1995; 72 no. 16:120.
Malpractice is not criminal; mistake is? *The National Law Journal*. July 28, 1997; podium:A17.
Will this internist's name finally be cleared? Gerald Einaugler. *Medical Economics*. March 10, 1997; 74 no. 5:50.

Criminal conviction upheld, appealed. *Medical Malpractice Law & Strategy*. April 1997; 14 no.6:1.

Doctors in court. *Capital Times (Madison, Wis.)* April 12, 1995; nation/world:2C.

Doctor's sentence delayed. *The New York Times*. March 18, 1995; sect 1:26.

Fein EB. Doctor in negligence case gets his sentence eased. *The New York Times*. June 28, 1997; sect 1:25.

Kriegel M. Law's treatment of doc is criminal. *Daily News (New York)*. May 21, 1997; news:19.

Malpractice or homicide? [editorial]. *The Washington Post*. April 18, 1995; editorial:A16.

McShane L.
Doctor spending weekends in jail: botched procedure in 1990. *The Record (Bergen Record Corp)*. May 18, 1997; news:A6.
Doctor's error leads to death, jail. March 16, 1995.

Medical errors becoming crimes. *Chattanooga Free Press*. April 12, 1995.

Moy DR. "Cover-up" allegation called unjustified [letter]. *New York Law Journal*. September 12, 1995; letters:2.

Nossiter A. A mistake, a rare prosecution, and a doctor is headed for jail. *The New York Times*. March 16, 1995; sect A:1.

Pines D. Doctor loses death-case appeal: negligence affirmed in treating patient. *New York Law Journal*. March 11, 1997:1.

Susman C. When doctors err. *The Palm Beach Post*. May 27, 1995; accent:1D.

Weiner SM, Reibman M. Doctors will get more immunity for mistakes [letter]. *The New York Times*. March 22, 1995; sect A:18.

A-7 (Colorado Nurses)

Newborn boy dies after improper shot. *The Washington Times*. October 25, 1996; part A, nation, American scene:A11.

Romano M.
Baby dies after injection: police, state probe death at St. Anthony Hospital North. *Rocky Mountain News*. October 24, 1996; local, ed F:5A.
Nurses fear indictments put profession on trial: charges in infant's accidental death are a first in Colorado. *Rocky Mountain News*. April 30, 1997; local, ed F:28A.

Metro digest: Adams County nurses trial ruling. *The Denver Post*. June 19, 1997; Denver & the West:B2.

Schrader A, Robinson M.
Baby's nurses face homicide charges. *The Denver Post*. April 29, 1997; sect A:1.
Healthy baby's tragic death at hospital being probed. *The Denver Post*. October 25, 1996; Denver & the West:B7.

UPI. Colo. newborn dies of improper injection. October 24, 1996.

Sourcebook: Workshop on Assembling the Scientific Basis for Progress on Patient Safety. Volume II

Section B. Synthetic views

Also from the popular press are two broader perspectives from reporters looking across multiple accidents and attempting to synthesize a more coherent picture of error and accident in general. The article by Belkin is the most recent and most detailed piece to date. Also included is a broadside by the news commentator Sidney Zion, father of Libby Zion. This response to a media briefing conducted by the National Patient Safety Foundation provides some suggestion of the charged atmosphere surrounding patient safety issues and may presage the sorts of difficulty confronting those who seek to develop and explore a more technically grounded view.

B-1

Belkin L. How can we save the next victim? *The New York Times Magazine*. June 15, 1997; sect6:28-33, 44, 50, 63, 66, 70.

Cohen M. How can we save the next victim? [letter]. *The New York Times*.

Ho L. How can we save the next victim? [letter]. *The New York Times*.

Meyer RH. How can we save the next victim? [letter]. *The New York Times*.

B-2

Medicine's hidden costs: lack of a reporting system permits repetition of mistakes [editorial]. *Pittsburgh Post-Gazette*. October 31, 1993; editorial:D2.

Twedt S.
Deadly hospital mistakes are doomed to be repeated. *Pittsburgh Post-Gazette*. October 24, 1993; national:A1.
Despite all his precautions, an error killed his mother. *Pittsburgh Post-Gazette*. October 28, 1993; national:A13.
Hospital computer reveals the future. *Pittsburgh Post-Gazette*. October 28, 1993; national:A13.
A hospital drug mix-up turns into fatal error. *Pittsburgh Post-Gazette*. October 24, 1993; national:A8.
Hospitals blind to each other's mistakes.*Pittsburgh Post-Gazette*. October 27, 1993; national:A1.
Hospitals to discuss prevention: regional forum will share ideas, methods. October 31, 1993; national:A1.
How two lost lives helped to reform a state. *Pittsburgh Post-Gazette*. October 27, 1993; national:A8.
Lessons hard to learn if state requires secrecy. *Pittsburgh Post-Gazette*. October 27, 1993; national:A8.
"My God! That must be what I gave." *Pittsburgh Post-Gazette*. October 25, 1993; national:A6.
Prescriptions fraught with danger. *Pittsburgh Post-Gazette*. October 25, 1993; national:A1.
The slow search for an antidote to health care's deadly problem. *Pittsburgh Post-Gazette*. October 28, 1993; national:A13.

Syringe confusion leaves a trail of deadly mix-ups. *Pittsburgh Post-Gazette*. October 26, 1993; national:A1.

Tragic errors strike baby 3 times. *Pittsburgh Post-Gazette*. October 25, 1993; national:A7.

B-3

Zion S. At long last, some docs 'fess up. *New York Daily News*. October 13, 1997; opinion.

Section C. Uncelebrated cases

These items represent areas where some other research base is available and are chosen because they highlight some of the factors believed to affect success and failure in the particular area.

Of particular note are the results of the Harvard Medical Practice Study. This seminal study has stimulated many to look harder at the factors affecting patient safety. The study has been widely quoted and its extrapolations have become the basis for many of the estimates of the "cost of error" and scope of the problem as well as the source of sometimes heated debate within the medical community.

C-1

Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991; 324:370-376.

Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard medical Practice Study II. *N Engl J Med*. 1991; 324:377-384

C-2

Stewart L, Way LW. Bile duct injury during laparoscopic cholecystectomy: factors that influence the results of treatment. *Arch Surg*. 1995; 130:1123-1128.

Way LW. Bile duct injury during laparoscopic cholecystectomy (editorial comment). *Annals of Surgery*. 1992; 215:195.

C-3

Dominguez CO, Flach JM, McKellar DP, Dunn M. Using videotaped cases to elicit perceptual expertise in laparoscopic surgery. *Proceedings: Third Annual Symposium on Human Interaction With Complex Systems: HICS '96*. Los Alamitos, Calif:IEEE Computer Society Press, 116-123

C-4

Cook RI, Wood DD, Howie MB, Morrow JC, Galsa DM. Unintentional delivery of vasoactive drugs with an electromechanical infusion device. *Journal of Cardiothoracic and Vascular Anesthesia*. 1992; 6(2): 238-244.

C-5

Bates DW, Leape LL. Pharmacies and prevention of potentially fatal drug interactions. *JAMA*. 1996; 275:1088

Cavuto NJ, Woosley RL. Sale M. Pharmacies and prevention of potentially fatal drug interactions. *JAMA*. 1996; 275:1088

Kraft KE, Dore FH. Computerized drug interaction programs: how reliable? *JAMA*. 1996; 275:1088.

C-6

Smith PJ, Galdes D, Fraser J, Miller T, Smith JW Jr, Svirbely JR, Blazina J, Kennedy M, Rudmann S, Thomas DL. Coping with the complexities of multiple-solution problems: a case study. *Int J Man-Machine Studies*. 1991; 35:429-453.

C-7

Guerlain S. Using the critiquing approach to cope with brittle expert systems. *Proceedings of the 1995 Annual Meeting of the Human Factors and Ergonomics Society*.

C-8

American Academy of Orthopaedic Surgeons.
Report of the AAOS Task Force on Wrong Site Surgery.
Advisory statement: wrong-site surgery. September, 1997.

Canadian Orthopaedic Association. Committee on Practice and Economics. Position paper on wrong sided surgery in orthopaedics. June 10, 1994.

C-9

Cook RI. Human error in the ICU: an episode. 1996.

West S, Walch K, Youngberg, BJ. (1994) Risk management program development. In: Youngberg BJ (ed). *The Risk Manager's Desk Reference*. Gaithersburg, MD: Aspen.

Addendum

Billings CE. Some hopes, and concerns, regarding medical event reporting systems: lessons from the NASA Aviation Safety Reporting System (ASRS). *Arch Pathol Lab Med*. 1998; 122(3): 214-215

Donchin Y, Gopher D, Badihi Y. Recording and analysis of human errors in the intensive care unit - mission possible?!

Donchin Y, Gopher D, Olin M, Badihi Y, Biesky M, Sprung CL, Pizov R, Cotev S. A look into the nature and causes of human errors in the intensive care unit. *Critical care medicine* 1995; 23 no. 2:294-300.

Fischhoff B. Communicate unto others... *Reliability Engineering and System Safety*. 1998.

Gaba D. Organizational structures, cultures, and system aspects of safety in tertiary health care: a comparison with other high-risk industries. (Submitted to the Robert Wood Johnson Foundation Investigator Awards in Health Policy grant program, July, 1996).

Goldschmidt G, Donchin Y. Breathing circuit disconnection during mechanical ventilation can be prevented! A human factors look into a system: report of a case and analysis.

Grabowski M, Roberts K. Risk mitigation in large-scale systems: lessons from high reliability organizations. *California Management Review*. Summer, 1997; 39 no. 4:152-162.

Leape LL. A systems analysis approach to medical error. *Journal of Evaluation in Clinical Practice*. 1997; 3 no.3:213-222.

Margolis H. Introduction: Why the public and the experts disagree on environmental issues. In *Dealing With Risk*. Chicago: The University of Chicago Press, 1996:1-5.

Rasmussen J. Risk management in a dynamic society: a modeling problem. *Safety Science*. In press.

Roberts KH, Libuser C. From Bhopal to banking: organizational design can mitigate risk. In *Organizational Dynamics*. American Management Association, 1993:15-26.

Tasca LL. *The social construction of human error* [dissertation]. Stony Brook: The State University of New York; 1990:iii, iv, vi, 23-60.

Index

National Patient Safety Foundation
at the AMA

515 North State Street, 8th Floor
Chicago, Illinois 60610

Telephone 312 464-4848
E-mail npsf@ama-assn.org
Web http://www.npsf.org