

6-5-2007

The Validation of a Methodology for Assessing the Impact of Hybrid Simulation Training in the Minimization of Adverse Outcomes in Surgery

Peter J. Fabri
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

Scholar Commons Citation

Fabri, Peter J., "The Validation of a Methodology for Assessing the Impact of Hybrid Simulation Training in the Minimization of Adverse Outcomes in Surgery" (2007). *USF Tampa Graduate Theses and Dissertations*.

<https://digitalcommons.usf.edu/etd/703>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

The Validation of a Methodology for Assessing the Impact of Hybrid Simulation
Training in the Minimization of Adverse Outcomes in Surgery

by

Peter J. Fabri, M.D.

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Jose Zayas-Castro, Ph.D.
Michael T. Brannick, Ph.D.
Robert P. Carnahan, Ph.D.
Kingsley Reeves, Ph.D.
Michael X. Weng, Ph.D.

Date of Approval:
June 5, 2007

Keywords: Graduate Medical Education, Human Error, Medical Error, Patient Safety,
Surgical Complications, Surgical Outcomes

© Copyright 2007, Peter J. Fabri, M.D.

Dedication

This manuscript is dedicated to all health care professionals who consider themselves “Health Care Engineers,” and who aspire to improve patient care by the application of systems and processes to the clinical practice of medicine.

Acknowledgements

I would like to acknowledge my wife, for providing support, encouragement, and editorial assistance; my children and my grandson for providing the stimulus to persevere; my staff, for helping to carry the load during the years of study; my colleagues for making it feasible for me to pursue this goal; the faculty and students of the Department of Industrial and Management Systems Engineering for accepting me into their ranks; the Department of Veterans Affairs for supporting me during my studies; the University of South Florida College of Medicine for providing me support, time, and collaboration; and the USF Department of Surgery faculty and residents for actively participating in the study design and implementation.

Table of Contents

List of Tables	iii
List of Figures	iv
List of Abbreviations, Initializations and Acronyms	v
Abstract	vii
1. Introduction	1
2. Literature Review	8
2.1 Literature on Error	8
2.2 Literature on Medical Error	13
2.3 Literature on Simulation	15
2.4 Literature on Graduate Medical Education	16
2.5 Literature on Classification of Error (Taxonomy)	17
2.6 Literature on Transfer of Training	18
2.7 Literature Summary	19
3. Methods	20
3.1 Classification and Scoring Systems	20
3.2 The Scoring Template	25
3.3 Validation	29
3.3.1 External Validation	29
3.3.2 Internal Validation	30
3.4 Case Review Methodology and Data Classification	31
3.5 Near-Miss Reporting	32
3.6 Preliminary (Formative) Data Review	33
3.7 Designing the Training Program	33
3.7.1 Developing the Videotaped Scenarios	35
3.7.2 Developing the Mannequin Based Simulation Training	36
3.8 Institutional Review	38

3.9 Data Analysis and Statistics	38
3.9.1 Reliability Analysis	39
3.9.2 Validity Assessment	39
3.9.2.1 Face Validity	39
3.9.2.2 Construct Validity	40
3.9.2.3 Criterion Validity	40
3.9.2.4 Content Validity	41
3.9.3 Data Classification Analysis	41
4. Results	43
4.1 External Template Validation	44
4.2 Internal Comparison	54
4.3 Video Evaluation	61
4.3.1 Video Script Review	61
4.3.2 Video Review	61
4.4 Analysis of Baseline Error Data	61
4.5 Development and Validation of the Training Program	67
4.6 Post-Training Data Analysis	70
5. Discussion	83
5.1 Analysis of Results	83
5.2 Future Plans	93
5.3 Conclusion	95
References	97
Bibliography	105
Appendices	108
Appendix 1 Letter to Expert Evaluators for External Validation of Scoring Template	109
Appendix 2 Mock-up of Scoring Template with Instructions	110
Appendix 3 Glossary of Terms Used in the Error Template and Instructions	111
Appendix 4 Instructions for Using the Error Evaluation Template	114
Appendix 5 Morbidity and Mortality Scenario 1 for External Validation	115
Appendix 6 Morbidity and Mortality Scenario 2 for External Validation	116
Appendix 7 Morbidity and Mortality Scenario 3 for External Validation	117
Appendix 8 Scenario Script: Operating Room	118
Appendix 9 Armchair Conference	120
Appendix 10 Follow-up Scenario Script: Without Errors	122
Appendix 11 Fact Sheet0 STAN, the METI Human Patient Simulator	124
Appendix 12 Simulation Script: Recovery Room	126
About the Author	End Page

List of Tables

Table 1.	Error Classification According to Reason	12
Table 2.	Years Since Completing Residency	44
Table 3.	Outcome Score for the Three Scenarios	45
Table 4.	Agreement in Error Classification in the Scenarios	47
Table 5.	Characterization of Error Types in Validation Scenarios	50
Table 6.	Characterization of All Errors Listed in Validation Scenarios	51
Table 7.	Scenario Data Analysis Weighted by Position in List	52
Table 8.	Agreement Between the Two Evaluators	55
Table 9.	Internal Evaluator Outcome Scores	55
Table 10.	Distribution of Error Types in Inter-rater Comparison	59
Table 11.	Monthly Procedures, Complications, and Errors During Baseline Period	63
Table 12.	Analysis of Characteristics of Individual Error Types	66
Table 13.	Analysis of Error Data by 6-Month Blocks	71
Table 14.	Analysis of Morbidity and Mortality Data by Month Over 18-Month Study Period	72
Table 15.	Regression Analysis of Principle Components Over Time	74
Table 16.	The Outcome Scores Before and After Training	79

List of Figures

Figure 1.	Error Types Identified in Scenarios	49
Figure 2.	Comparison of Error Scores (1-5) Between Evaluators 1 and 2	56
Figure 3.	Residual Analysis of Error and Index Error	57
Figure 4.	Graphical Comparison of Errors Identified by Internal Raters	60
Figure 5.	Graph of Number of Errors per Month During Baseline	62
Figure 6.	Frequency Distribution of Error Types During Baseline	64
Figure 7.	Surgical Procedures, Complications, and Errors by Month	75
Figure 8.	Errors as Percent of Procedures and Percent of Complications	76
Figure 9.	Index Error as Percent of Procedures	77
Figure 10.	Index Error as Percent of Complications	78
Figure 11.	Trend Analysis for Complications, Errors, Index Errors	81
Figure 12.	Error Template as Sent to Evaluators	110

List of Abbreviations, Initializations and Acronyms

AAMC	Association of American Medical Colleges
ACGME	Accreditation Council for Graduate Medical Education
ACLS	Advanced Cardiac Life Support
ACS	American College of Surgeons
AMA	American Medical Association
ANOVA	Analysis of Variance
ATLS	Advanced Trauma Life Support
BP	Blood Pressure
CC	Cubic Centimeter
CT	Computerized Tomography, CAT Scan
DVT	Deep Venous Thrombosis
EKG	Electrocardiogram
FST	Frequency, Severity, Timing
GI	Gastrointestinal
GME	Graduate Medical Education
HR	Heart Rate
ICU	Intensive Care Unit
JCAHO	Joint Commission on Accreditation of Healthcare Organizations
KS	Kolmogorov Smirnov test of distributions
LOS	Length of Stay
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
METI	Medical Education Technologies, Inc.
MG	Milligram
MI	Myocardial Infarction (Heart Attack)
ML	Milliliter
MSD	Mean Standard Deviation
OR	Operating room

PACU	Post Anesthesia Care Unit (Recovery Room)
POD	Post Operative Day
SD	Standard Deviation
SEM	Standard Error of the Mean
STAN	STAN is a human physiologic simulator
TPN	Total Parenteral Nutrition
UCLA	University of California, Los Angeles
USF	University of South Florida
VA	Veterans Affairs Hospital, Department of Veterans Affairs
WBC	White Blood Count

The Validation of a Methodology for Assessing the Impact of Hybrid Simulation Training in the Minimization of Adverse Outcomes in Surgery

Peter J. Fabri, M.D.

ABSTRACT

The Institute of Medicine report “To Err is Human,” released in late 1999, raised the issue of human error in medicine to a new level of attention. This study examines the frequency, severity, and type (FST) of errors associated with postoperative surgical complications at a tertiary care, university-based medical center, addressing the intersection of three domains: patient safety, graduate medical education, and simulation-based training. The study develops and validates a classification system for medical error that is specific to surgery, affirming reliability internally and externally. Baseline data on the FST of errors is collected over a 12-month period. A hybrid, simulation based training session is developed, validated, and applied to a cohort of surgical residents, focusing on the three most common types of errors identified from pilot data, namely judgment error, incomplete understanding of the problem, and inattention to detail, all human factor errors. The impact of the training is evaluated by measuring the FST of errors occurring during the 6-month period following the training sessions. The study demonstrates that there is a continuous decrement in the incidence of postoperative complications and a proportional decrease in error, which starts at the beginning of the

baseline data collection and continues linearly throughout the 12 baseline months and subsequent 6 post-training months. There is no additional decrement in the rate of change following training, and no change in the rate of the index errors following the training. This study suggests that surgical error is frequent (>2%) and principally due to human factors rather than systems or communication. This study demonstrates that creating an environment where residents are continuously involved in identifying and characterizing errors results in a significant and sustained decrease in postoperative complications and the errors specifically associated with them. Contrary to expectations, a validated, well-designed, active-learning training module does not result in an additional identifiable improvement in patient outcome or in the incidence of index errors. These results are at variance with many recent studies addressing medical error and, if verified by additional studies, challenge several strongly held ideas related to patient safety training.

1. Introduction

Since the publication of the Institute of Medicine report *To Err Is Human* in 2000,^{1(p25)} patient safety has become a high priority in U.S. healthcare and in medical education.²⁻¹⁰ Earlier work by Leape,¹¹⁻¹³ Gaba,¹⁴ and many others¹⁵⁻²¹ had identified the importance of active and structured interventions to decrease the risk of error at all levels of medical practice. In perhaps the earliest work addressing the sociology of surgery, Bosk reported how an academic department of surgery addressed error within its culture.²² Most notably, these studies have identified that the two largest areas of risk in healthcare are errors¹⁷ related to surgery and to medication errors. Furthermore, many academic surgeons have felt that having residents “in house” 24 hours per day improves patient care and enhances patient safety, particularly as healthcare has become increasingly complex. The residents were believed to be able to “work around” many of the system problems that have been introduced by modern technology. Said differently, resident physicians, being trainees, are part of the “cause” of medical error, but at the same time may have been an important component of the “safety net” that identifies and minimizes error. The purpose of this study is to determine if a structured, simulation-based training program, aimed at surgical residents, can improve actual measured patient outcomes.

Traditional studies of error analysis^{23, 24} have identified three broad approaches to minimizing error: redesigning systems to “eliminate” the risk of error, creating methods to “highlight” an error so that it is immediately recognized, and developing approaches to mitigate the adverse consequences of errors when they happen. Numerous studies have addressed methods of redesigning medical systems to make them safer.²⁵⁻²⁸ Like the nuclear power and aviation industries, which have served as the principal “models” for error reduction in medicine,^{29, 30} high technology areas in healthcare (such as radiation therapy and anesthesiology) may lend themselves well to reengineering in order to mitigate error risk. To the degree that surgery is inseparable from anesthesia, surgery too may be improved by efforts to reduce system error. But when surgical errors occur in the operating room, it is rare that the technology itself is the cause of the problem. Rather, it is the human decision process or human manipulation of the technology that leads to difficulties.³¹ It is unclear how much of the error risk of surgery can be reduced by redesigning systems or reengineering.

Team training and crew resource management,³² two terms used to describe structured training of dedicated teams in the aviation industry, have been applied to the operating room environment³³⁻³⁶ and critical care areas^{37, 38}. This certainly has direct application to the “system” errors which occur in the operating room³⁹. Many “catastrophic” events that occasionally occur in the operating room, such as wrong side surgery, operating on the wrong patient, or leaving a sponge or instrument within a patient can be “eliminated” by improving the teamwork among individuals. Similarly, no one can argue that improving communication and teamwork in the operating room is not

desirable. But it has yet to be demonstrated that this will decrease surgical error or improve the outcome of surgery.

Some authors have suggested that it is not possible to eliminate the risk of error when the principle component is human-to-human interaction^{40(p198)}. This situation, individual human-to-human interaction, would appear to be the case in much of surgical care, which raises the possibility that error relating to the thought processes, decision processes, and technical processes of the surgeon may be an unavoidable consequence. Although the group of individuals who work together within an operating room is often referred to as the “surgical team,” most authors⁴¹ admit that this is a haphazard team at best, and perhaps might actually be considered three or four separate “teams” that often work at cross purposes to each other. Increasing communication among these separate teams will improve surgical care. But it will not address the category of errors which originate from the decisions and technical activities of the individual surgeon.

It is critically important to define the types of errors that occur in the operating room, their frequency, and their severity. In addition, it is important to attempt to identify the cause of the errors, to define whether they are amenable to reengineering or training. Many error studies have identified a very high percentage (often greater than 80%) of errors being due to human factors. Some authors have suggested that this is just the “sharp end”⁴² of the problem, where it actually occurs, but that in fact there is a cascade of events beginning with system errors on the “blunt end” upstream in the process. Is this the case with surgical error? To date, no study has attempted to answer this question in relation to surgical errors.

Finally, many investigations in human error have suggested that the presence of humans even in highly technical processes is essential, because it is not possible to design out all potential errors in highly complex systems^{40(p180)}. The creativity and adaptability of the human participant allows new and previously unrecognized problems to be addressed in real time. However, given the potential for a large number of such problem solving experiences in medicine (and in surgery) the potential for error is probably an intrinsic component. Some have even argued that it is surprising that error does not occur more frequently. If this is true, what is the role of surgical trainees in minimizing error? As surgical care has become more acute, more complex, and more technologically sophisticated, have residents been the “glue” that holds a loosely coupled, highly complex system together? If residents have served the role of identifying and minimizing the consequences of errors that are “built in” to our systems, is it possible that training residents to recognize and respond to error will improve the measurable outcome of surgical care?

This study will address the intersection of three specific domains within the broad field of medicine: error in surgery, the use of simulated training, and surgical residency. This study will develop and apply a structured training module to introduce residents to the most common types of error, as measured during the preliminary phase of data collection. Creating this model will require the prior development of a classification system for error in the surgical environment, measurement of the frequency and types of error and their consequences, and measurement of the patient outcomes that follow. After design and application of the training program, this research will assess the impact of the training program on errors occurring after the cohort of surgical residents has been

“trained” by repeating the measurement of the frequency, types, and severity of errors in the post-training period and the patient outcomes that occur.

The specific goals are:

1. to develop a satisfactory classification for medical error within the discipline of surgery,
2. to validate the instruments developed to measure medical error,
3. to define the baseline incidence of error occurring in patients who have complications of surgery,
4. to define the most common types of error associated with surgical complications,
5. to measure the impact of an error in the actual outcome of patients,
6. to develop a reliable training program that uses simulation to train surgical residents to be aware of and avoid medical errors, and
7. to determine if a simulation-based training program delivered to an entire cohort of surgical residents can result in a measurable change in patient outcome.

The questions (research hypotheses) which will be tested are:

1. Is it possible to create a valid classification system for errors occurring in surgery.
2. Can this classification system demonstrate reliability when used by practicing academic physicians.
3. Can we determine a baseline incidence of the types of common errors.
4. Can we reliably identify the most common types of errors in an academic surgical setting.
5. Can we measure the impact of error on the actual outcomes of patients.
6. Can we develop and validate a simulation-based training module for resident physicians.
7. Can a validated training module improve surgical outcomes after a cohort of residents are trained.

It is important to clarify what this research will NOT address.

1. It will not consider errors outside of the field of surgery in an academic medical center environment. By design, this study only addresses reported “complications” occurring after surgical procedures in an academic medical center.
2. It will not address redesigning systems and processes in the hope of eliminating error. The only designed intervention is a training session for

residents. There will be no attempt to address underlying or latent problems that might be amenable to redesign or reengineering.

3. It will not address “near-miss” or latent errors that have not already led to adverse patient outcome. The original design of this study included measuring “near-miss” events concurrently, but this was very unsuccessful, as described later in this document.

4. It will not address team training or training in groups. Although the “team” is assumed in addressing a surgical event, surgical residents are trained individually. The residents are working in a simulated clinical environment with a nurse, but the nurse is an “actor” and not a “trainee.”

5. It will not address the immediate effects of the training experience, but rather will focus on the downstream impact on patient outcome. Many studies measure short term, direct effects of simulation-based training, such as performance on a technical procedure or a multiple choice text.

This study only measures far transfer as measured by patient outcomes at a later time.

2. Literature Review

This study addresses three primary content domains and several subsidiary areas. Each has an extensive and largely non-overlapping literature base. Each of these is reviewed individually to provide the relevant background information to inform the reader, and then briefly summarized in the aggregate.

2.1 Literature on Error

This literature review was conducted by first performing a National Library of Medicine search from 1966 to the present using the keywords “error” and “patient safety.” Subsidiary searches were conducted using the term “graduate medical education” and the term “simulation.” All references included in the website of the Cognitive Technologies Laboratory (kindly provided by Dr. Richard Cook) were reviewed. Searches were performed using the State of Florida library system on the terms “error”, “human error”, “medical error”, and “patient safety.” All books identified in the search were read and critiqued in detail. The bibliographies of all items identified by the above searches were extensively reviewed to identify important prior work to avoid omission. While great effort was made to be inclusive, this work addresses several major, overlapping (and some non-overlapping) domains and an immense literature base. It is therefore possible that something has been overlooked.

Man has been concerned with the subject of error at least since the time of Plato. In *The Republic*⁴³, Plato addresses the “shadow on the wall” and its lack of total harmony with the ideal. Later, in the middle ages, Roger Bacon⁴⁴ (13th century) referred to “offendicula” or impediments of thought and produced perhaps the first “classification system” of error, recognizing four “obstacles” to inquiry: overweight of authority, slavery to custom, dominance of popular opinion, and the concealment of ignorance by pretense of knowledge. Centuries later, Francis Bacon⁴⁴ also addressed human error from a philosophical basis, offering a more “advanced” classification system. He described four “idols”: the idol of the Tribe--the insistence that Nature is as the human mind would have it; the idol of the Cave--limitation of outlook making men non-understanding and intolerant of others’ ways; the idol of the Forum--the undue deference to public opinion; and the idol of the Theater--the conditions of the scholastic contingent directed at authority, tradition, vested interests and false notions. It is interesting to reflect back on these descriptions as we attempt to understand error in the modern world. Each of these authors of antiquity (and others) focused on error as a component of the philosophy of the mind, or metaphysics.

The 20th century saw the introduction of in-depth and structured approaches to human error in the new fields of cognitive psychology, neurophysiology and psychiatry. These domains dominated the study of error from the early 1900’s through the 1960’s. Even Freud⁴⁵ described a concept of human error as a component of his theories of psychoanalysis, leading to the commonly used phrase “Freudian slip.” Codman, in the early 1900’s, introduced perhaps the first structured classification system of surgical error^{30, 46}. He reviewed 123 surgical errors in 337 patients and proposed a set of seven

types of error: “Error due to lack of technical knowledge or skill, error due to lack of surgical judgment, error due to lack of care or equipment, error due to lack of diagnostic skill, the patient’s unconquerable disease, the patient’s refusal of treatment, and the calamities of surgery or those accidents and complications over which we have no known control.” This list is surprisingly similar (with changes in language use) to the results of the present investigation. In *Managing Your Mind*⁴⁷, Jastrow (1931) presented a new, “modernized” classification of error which defined three subjective and three objective idols: the idol of the Self- the projection of the subjective upon Nature; the idol of the Thrill- the favoring of the romantic and dramatic; the idol of the Web- the spinning of imaginative data; the idol of the Mass- undue deference to popular opinion; the idol of the Mold- restriction to one’s own class-cast of mind and outlook; and the idol of the Cult- bondage to dogma and ‘isms. Later (1936) Jastrow wrote in *The Story of Human Error* an entire chapter entitled “Error in Medicine”, which describes the historical progress of medical thought, concluding with the provocative thought. “Today not only the physician must know the pitfalls that are the errors of medicine, but the patient must also know them if he too will avoid them.” Although these writers were describing a much more concrete concept of error and error classification than their historical antecedents, their works still largely address error as philosophical and metaphysical in nature.

Studies in the 1960’s and 1970’s were heavily motivated by new advances in behavioral psychology. It wasn’t until the late 70’s, however, that interest in nuclear reactor risk resulted in the development of systematic and mathematically based studies of human reliability analysis, systems engineering, and human factors research. As

computer power increased, concepts that had been developed in the fields of artificial intelligence and neural networks were applied to system engineering, and error analysis became more analytical and process driven. A specific formulation based on a rule-based computer system was proposed by Newell and Simon⁴⁸ known as the “General Problem Solver.” Rouse⁴⁹, Rasmussen⁵⁰, and Reason⁴⁰, among others, advanced this concept of rule-based systems and independently developed coherent concepts of human error, which built on each other. Rouse described a problem-solving approach based on an “if (situation) ...then (action)” approach to human error. This is often referred to simply as a “rule-based” system.

Rasmussen⁵¹ extended this concept into a “skill-rule-knowledge” classification system. This concept forms the kernel of our modern understanding of error. He categorized error as being skill-based, rule-based, or knowledge-based. He formulated a baseline, problem-solving method using cognitive knowledge, a method used by young or new learners, but also reactivated in more advanced learners when solving new or complex problems. In this situation, the individual explores a large, internal knowledge base and develops a unique solution to the problem (knowledge-based system). As individuals develop a sense of recurrent patterns within problems, they identify a set of rules, some formal, some informal, which can be applied in the same or similar settings (rule-based system). When individuals become very familiar with the rules and can identify when the rules need to be modified to fit novel applications, problem solving becomes dominated by utilization of stored patterns of preprogrammed internal instructions which are organized within a time-space domain (skill-based system).

Error^{40(p9)} was thus classified (Table 1) as depending on a violation of a corpus of knowledge, breaking a specific rule, or the failure of a highly developed skill.

Table 1. Error Classification According to Reason^{40(p12)}

Type of Error	Classification	Timing
Knowledge based	Knowledge based mistake	Evaluation/Planning
Rule based	Rule based mistake	Evaluation/Planning
Skill based	Lapse (storage) Slip (execution)	Execution

More recently, Senders¹⁶ brought together 22 internationally recognized error experts for a workshop on human error (The Bellagio Conference). They reviewed and discussed the current classification systems, theories of error, and approaches to error mitigation and concluded that there is no single, useful classification system for error. They argue that classification of error is uniquely related to the purpose of the research, and that no single theoretical model of human error will either explain all error or eliminate its occurrence^{40(p10)}. Wallace⁵² essentially repeated this concept of the absence of a unifying classification system. He did, however, emphasize the importance of determining the reliability of any classification system, citing that a minimum of 70% reliability is essential to analyze data in a meaningful and unbiased manner. However, a noted characteristic of the existing literature on human error is the absence of reliability

testing or hypothesis testing. This study will address these absences by applying statistical design and rigorous data analysis.

2.2 Literature on Medical Error

The first modern, “scientific” studies to call attention to the high incidence of error in the practice of medicine emerged from Leape’s work with the Harvard Practice Group Study in 1991^{53, 54}. Bogner^{17, 55} (1994), from the U.S. Food and Drug Administration, published an extensive review of the published work in the field of medical error. This work began with the concepts of Reason (which was entirely non-medical) and incorporated ideas developed by Leape, Gaba, Woods, and others. Gaba¹⁴, focused principally on error within the field of anesthesiology. He presented a thorough description of system re-engineering for safety and explained the role of team training in minimizing error. This laid the scientific ground work for the first major “call to arms,” the now famous Institute of Medicine report in 2000 entitled *To Err is Human*. This document, extrapolating from the data from the Harvard Practice Group Study, estimated that perhaps more people die as a consequence of medical error than from automobile accidents. *To Err Is Human*, in part because it represented the views of such a highly recognized and credible group (The Institute of Medicine,) could not be ignored. It triggered an avalanche of interest in the subject of error in medicine, both in the scientific community and in the lay press. Since then, numerous articles have reinforced the costs and implications of medical error. Recently, many authors have attempted to develop classification systems for medical error, based on the early works of Reason^{23, 40, 56, 57} and more recently summarized by Zhao and Olivera⁵⁸. Most of the published studies of

medical error relate to the office-based practice of family medicine⁵⁹⁻⁶² and the errors are predominantly clerical, administrative, and prescription related, while the preponderance of studies conducted in hospitals look at medication errors (prescribing, dispensing, administering). Other studies focus on very specific technical procedures, the complications and learning curves associated with them, and the role of training on short-term and directly measurable technical outcomes. There are no current studies which address a classification system for error in the broad field of general surgery or methods to address the role of training in minimizing the impact of medical error in surgery.

During the same era in which medical error was being highlighted, attention was drawn to the role of residents in adverse patient outcomes. Although much of the published literature has focused on the role of work hours and fatigue on resident performance, the critically important role of residents in both causing and preventing errors has also been emphasized. The introduction of the “Outcome Project”^{63, 64} by the Accreditation Council for Graduate Medical Education in 2001 was intended to transform resident training from an apprenticeship model to a competency-based model and thus assure that residents are fully trained and competent to practice medicine. Of perhaps more importance, however, the “Outcome Project” formalized the already growing emphasis of supervision of residents in the conduct of patient care. Simultaneously American society was no longer comfortable having residents “practice” on human patients. The combination of the recognition of the incidence of error, the hazards associated with trainees and young physicians, and the need to find new ways to train young physicians contributed substantially to the current emphasis on the development of high-fidelity simulation techniques in healthcare, and particularly in surgery.

2.3 Literature on Simulation

The earliest formal medical simulation devices were used to train students to listen to heart sounds (Harvey). More recently, mannequin-based simulators⁶⁵ have been used in training healthcare professionals in the technical aspects of advanced cardiac life support (ACLS) and cardiopulmonary resuscitation (CPR). Unlike the aviation industry, to which medical training is currently being compared and which has had high-fidelity simulators for decades, the medical community did not accept the importance of simulation in physician training until the public recognized the high cost of complications from laparoscopic cholecystectomy⁶⁶. Several simulation devices have since appeared and have been formally tested and validated⁶⁷⁻⁷⁰. Recent studies have addressed the role of technique simulators (e.g. laparoscopic surgery) in shortening the time for learning and perhaps decreasing technical complications⁷¹⁻⁷⁵. Resnick^{76, 77} et al have developed and evaluated a set of evaluation processes to assess resident performance. As reported by Gaba^{14, 78}, simulation has become a necessary tool in the training of anesthesiologists. Simulators of sufficient fidelity have been developed to allow realistic demonstrations of physiologic changes as a result of anesthetic agents, pharmacologic agents, and other acute physiologic alterations which can occur in an anesthetized patient undergoing an operation⁷⁹. To date, most of the published work in the area of simulation in surgery has focused on technical simulators⁸⁰. There is an absence of work in the use of simulation in training physicians to deal with errors which occur in surgical decisions and surgical practice⁸¹.

A number of investigators have looked at medical error with an effort to quantitate its incidence, define a classification system, and develop methods of prevention. These have largely focused on the outpatient practice of family medicine. In this setting, most of the errors appear to be related to scheduling, communication, and prescribing, rather than errors that are directly linked to interventions, as would be the case in studying surgical errors.

2.4 Literature on Graduate Medical Education

There is an extensive literature on the subject of Graduate Medical Education⁸², including a focused journal entitled “Academic Medicine.” Although there have been numerous articles on the use of simulation in resident education⁸³, including several in the field of surgery, there do not seem to be any published studies that look at patient outcome as a consequence of training in general or simulation-based training in particular. In 2001, the Accreditation Council on Graduate Medical Education (ACGME) introduced a new paradigm for the education and training of residents, the “Outcome Project^{63, 64}.” This ambitious undertaking focused initially on the control of the excessive numbers of work hours served per week by the typical resident, and which presumably was related to error and adverse patient outcomes. Studies on sleep deprivation, fatigue, and stress documented the adverse effect of these factors on patient safety. However, a more subtle yet critically important aspect of the “Outcome Project” has emerged over the past five years, as the program has been phased in. Traditionally, resident training has been a sort of “apprenticeship” for the advanced training of a trained “general physician.” As medicine has become more complex, it has become clear that an

individual is not “capable” at the completion of medical school plus an internship, and that the residency is best considered an extension of medical education and training. Concepts from the field of cognitive psychology and learning theory have been introduced into the medical education lexicon to address the process and methods involved in the progressive acquisition of knowledge, skills, and attitudes. Dreyfus and Dreyfus⁸⁴ describe the evolution of competence from “novice”, to “advanced beginner”, through “competent”, “proficient”, and “expert”. Competency is defined as the ability to execute a task or process independently and without supervision. Residents, by definition, are not fully competent, which implies that there is an increased risk of error associated with activities performed by residents. Inherent in the concept of a progressive, competency-based residency curriculum is the consideration that structured, systematic programs, formally defined in written goals and objectives, will focus and enhance the acquisition of competence in a progressive fashion. Patient safety is one of the “core competencies” that must be addressed within the formal curriculum of residency. Recent studies have begun to examine issues of patient safety within the education program of residency training.⁸⁵⁻⁸⁹

2.5 Literature on Classification of Error (Taxonomy)

Vincent³⁰ provides a thorough analysis of the history of error analysis in medicine and a context for further study in optimizing patient safety. Kopec summarizes the history of error classification and assesses several classification systems for medical error assessment.⁹⁰ Several authors⁹¹⁻¹⁰⁰ have focused on the development and use of classification systems for error and their limitations.¹⁰¹ Studies have also examined the

“chain” of events that make up a cascade, leading ultimately to error.^{20, 102-104} A detailed and comprehensive classification system of error has been developed by the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO)¹⁰⁵. Unfortunately, it is so comprehensive and complex that it requires a fully trained medical record coding specialist with supplementary training to be useful. Other classification systems have been “purpose specific” and do not lend themselves to use in other fields or circumstances. Wallace⁵² and others, in reviewing the state of the art in error classification have concluded that a single, all-purpose classification system (although one has been developed by the JCAHO) is not likely to be suitable for the “real-time” analysis of errors in healthcare. Wallace goes further to describe the ideal qualities of a classification system, specifically that it must be mutually exclusive and exhaustive (MEE) and that it must be reliable (reproducible) and valid (accurate). This review of the published literature in medical errors confirms the remarkable diversity of classification systems, but more importantly, highlights the stark absence of credible validation of the classification system or the impact of classification on a measurable outcome.

2.6 Literature on Transfer of Training

Much of the modern psychology literature (as well as works in the fields of education, communication, and engineering) emphasizes the importance of transfer of training. According to Noe¹⁰⁶, “Transfer of training refers to trainees effectively and continually applying what they learned in training (knowledge, skills, behaviors, cognitive strategies) to their jobs.” He further describes in detail the currently accepted model of the transfer process, which includes generalization of training to the job and

maintenance of learned material. Important antecedent work by Barnett¹⁰⁷, Ford¹⁰⁸, Tracey et al¹⁰⁹, and Wexley and Baldwin¹¹⁰ provides the scientific background for a more comprehensive analysis of transfer and its evaluation than will be addressed in this study. However, the idea of far transfer is critical to the analysis of this research—outcomes that occur in a different setting, at a different time, and not identical with the trained tasks. The published work and formal coursework by Brannick and Levine¹¹¹ assisted in the design of the training session and the outcome measurements used in this study.

2.7 Literature Summary

This study utilizes an expansive corpus of literature in the areas of human error, error in medicine, and transfer of training, plus an additional extensive literature on the subject of graduate medical education in general and the “Outcome Project” of the Accreditation Council for Graduate Medical Education (ACGME) in particular. Since the actual work performed in this study is in itself novel, the literature serves as an important background and platform for the investigations. Thus, the specific purpose of this study is to examine carefully the intersection of patient safety, graduate medical education, and the use of simulation. More specifically, this study attempts to identify whether the use of validated, simulation-based training can decrease the incidence, type and severity of medical errors in the actual practice of medicine. The study brings together the expertise of a large number of individuals and entities to address this question. Because this question has never been asked, a major component of this study is the development and validation of the instruments to measure error in surgery, and to quantify the baseline incidence.

3. Methods

An extensive review of the published literature was done in May 2005 for the purpose of defining the status of classification of medical errors in surgery. The review showed that most studies on medical error were in the primary care area (e.g. internal medicine, family medicine, pediatrics) and primarily addressed scheduling and communication errors. Prior work had identified a general taxonomy of error (organizational, mechanical, human factor), and this was used for the higher level classification. Since nothing has been published to identify causes of error within surgery, a list was generated from this researcher's prior experience of 30 years of weekly surgical morbidity and mortality conferences at two universities (Ohio State University; University of South Florida) combined with an analysis of the existing literature, previously cited.

3.1 Classification and Scoring Systems

Prompted by the work of Senders¹⁶ and of Wallace⁵², which affirms that there is no single, standard, acceptable taxonomy for the classification of error, this study developed a new classification system that was derived from the strongest published work and tailored to the common vocabulary of an academic surgeon. This classification system presumes that a complication has occurred, assesses whether an error was involved, determines the characteristics of the error, establishes a semi-quantitative score

for patient outcome, estimates the degree to which the error produced the complication, and classifies the type of error which occurred.

Error Presence. Predicated on the knowledge that not all adverse events are derivative of an error, the first criterion for the evaluator is the binary decision that there was or was not an error.

Patient Outcome. This study develops an outcome score that was patient centric, using a 5 point discrete classification scale, modified from the disability classification of the National Association of Insurance Commissioners severity of injury scale. The scores used in this study are

1 = If there was no adverse effect;

2 = If there was an identifiable adverse effect, which increased length of stay, but there was no discernible harm to the patient;

3 = If there was clear harm to the patient, but the patient either did or would be expected to fully recover with only temporary disability;

4 = If there was clear harm to the patient, the patient is expected to survive and improve, but with a permanent disability,

5 = If the patient died.

The outcome score is determined independent of the presence or type of error. If there was no error, the evaluation is terminated at this point, providing only a descriptor of the adverse event, the selection of “no error,” and the outcome score.

Error Characteristics. For those cases where the evaluator determines that there has been an error, the first question is if the error occurred during planning or during execution. Then, is the error a slip (correct action was planned but it was not performed correctly) or a mistake (incorrect action planned). Mistakes are not further subdivided into rule-based and knowledge-based, because of the difficulty of interpreting the “intent” and thought process of the “actors” at the time. This distinction will be addressed further in the subclassification of types of error.

Types of Error. The classification of types of error commenced with the three broad categories of error: organizational or system error, technical or mechanical error, and human error or human factors error. Since the term “technical error” in surgery has a different meaning than in industry (a technical error in surgery is in fact a human error), the term “mechanical error” is used to represent a situation where equipment malfunctioned. Since the most common errors evaluated were thought to be manifestations of human error (as described in most of the publications previously cited), and the focus of this research is on training humans, there was no further attempt to subclassify organizational and mechanical errors.

Human error, the specific focus of this exploration, is further sub-classified. This classification system began with Reason’s three categories: knowledge-based mistakes, rule-based mistakes, and skill-based slips and lapses, using language more familiar to a

surgeon. In considering knowledge-based mistakes, the phrase “incomplete understanding of the problem” is used as a non-judgmental descriptor. For rule-based mistakes, “judgment error” was substituted. In surgery, and in particular in mandated reviews of surgical complications, this concept appears to be widely understood and accepted. Skill-based errors (surgery being a discipline heavily dependent on the application of very advanced skills) were further subdivided into carelessness/inattention to detail (to address the notion of lapse) and technical error—in the surgical sense (to address the concept of slip). The entire list of error types is presented below:

Types of Medical Errors as Used in This Study

1. equipment failure (mechanical error)
2. health system error (organizational error)
3. incomplete understanding of problem
4. failure to use established protocol
5. carelessness/carelessness/inattention to detail
6. error in diagnosis
7. communication error
8. judgment error
9. delay error

10. error of omission

11. technique error

Error types are referred throughout this text by the coding numbers used in the computer reporting system, which are two digit adaptations of the listed error types, preceded by a “2.” Thus equipment failure is ErrorID 21.

Finally, to define a link between the error analysis and the current classification of core competencies in resident education, the evaluator was asked to identify which of the GME core competencies (Medical Knowledge, Patient Care, Communication, Professionalism, Practice Based Learning and Improvement, Systems Based Practice) were involved in the error/complication process.

This analysis of error taxonomy was reviewed with 6 senior academic surgeons at the University of South Florida and was modified based on their input. The total number of years of academic surgical experience among the group of individuals was in excess of 150. Although all are currently members of the faculty at a single university (University of South Florida), their individual origin and experience is from a large number of different universities with significantly different backgrounds. The extensive and diverse background of these individuals provided a broadly based initial draft classification. In order to enhance the preliminary “face validity” of the classification system, more extensive evaluation was achieved by means of a formal presentation of this classification system and the subsequent scoring template to the combined surgical faculty and residents at surgical grand rounds. Feedback from the aggregate department membership

was incorporated into the classification and the scoring template. Iterative changes were made in the template to enhance clarity, incorporating recommendations of the surgical faculty until the document reflected the collective understanding of the surgical faculty regarding error in surgery. Input of the residents was critical in improving the “user-friendliness” of the actual instrument used in the planned weekly complication assessment process.

This preliminary classification list was sent externally to 8 individuals who were selected on the basis of being nationally recognized academic surgical educators who had published in the field of patient safety, as obtained by doing a literature search of the National Library of Medicine. These individuals were each asked (Appendix 1) to respond anonymously, defining which of the items on the list were characteristics of medical error in surgery. They were also given the opportunity to provide additional error types to the list. A second request was sent to all 8 individuals to enhance return. A total of 4 individuals responded. There was uniform agreement with the classification system and comments of strong support from the respondents.

3.2 The Scoring Template

The higher level classification (organizational, mechanical, human) based on the published literature on error was combined with the broader error list, verified by the initial survey of surgical safety experts into a single scoring template, which would be used as the survey instrument throughout the remainder of the study. This template was reviewed with two experts in measurement (JK, College of Education; SN, College of Nursing) as non-surgical experts to assure that the document was easily understandable

without need for “insider information.” These “experts” provided critical insight into the need for and methods of establishing reliability and validity.

The final template (Appendix 2) included the following fields, each of which was described in both a glossary of terms (Appendix 3) and the instructions (Appendix 4):

A. *Patient ID*—an open field to allow a numerical sequence of cases, for example 1,2,3,4,5 at a presentation conference to affirm the order of the cases or the case scenario (A,B,C) during validation. Any inadvertent actual identification other than an ordinal was removed upon receipt.

B. *Reviewer*—an open field where a reviewer could place initials when paired comparisons were being performed. Otherwise, anything in this field was deleted.

C. *Score*—an interval scale for scoring the severity of complication into five levels

5 = The patient died during this episode of care

4 = The patient recovered but with a significant permanent disability

3 = The patient recovered but with a significant temporary disability

2 = The patient recovered without disability but with prolongation of hospitalization

1 = The patient had a “complication” but without any impact on the patient’s outcome.

D. *Complication description*—3 -5 word, open text “key words” describing the complication

E. *Medical Error*—a binary field (Yes or No) describing whether the reviewer believed an error had occurred

F. *Evaluation*—a binary field (Yes or No) describing whether an error occurred during patient evaluation

G. *Execution*—a binary field (Yes or No) describing whether an error occurred during patient treatment

H. *Slip*—a binary field (Yes or No) describing whether the error was caused by doing the appropriate thing but not doing it correctly.

I. *Mistake*—a binary field (Yes or No) describing whether the error was caused by doing something inappropriate (This field did not distinguish two subtypes—rule-based mistake and cognitive-mistake—as it was thought unlikely that academic surgeons would be able to consistently make this differentiation).

J. *Types of Error*—a classification list of the types of errors (see above) obtained from the preliminary validation phase of the study. The classification recognized the previously published division of types of error into organizational, technical, and human factor. Because these terms can have ambiguous meaning in clinical surgery, they were renamed “equipment failure (mechanical error), health system error (organizational error), and a comprehensive list of subcategories of human

factor error to represent various categories of human factor error. Additional open spaces were included to allow individuals to add free text.

K. *ACGME Core Competency*—a listing of the six ACGME core competencies to determine which competency seemed to be associated with the error. The actual wording of the competencies was altered slightly from the ACGME format for simplicity

1. Delivery of Patient Care
2. Insufficient or Inaccurate Medical Knowledge
3. Adult Learning; Problem Solving
4. Interpersonal and Communication Skills
5. Professional Behavior
6. Knowing the Healthcare System

To assure consistency in understanding/interpretation, a glossary defining the terms used in the scoring template and instructions was prepared. Where possible, definitions were taken from the published literature (Appendix 3). Evaluators were sent a detailed instruction sheet. (Appendix4). The instruction sheet was produced to guide individuals in the completion of the scoring template.

3.3 Validation

3.3.1 External Validation. To further validate the instrument, three surgical scenarios were created. (Appendices 5, 6, 7) Each scenario included sufficient information for an expert to be able to understand the clinical case and its implications, and sufficient information to be able to provide an opinion on the nature of the types of error which had occurred. The three scenarios, the scoring template that resulted from the previously described validation, the detailed glossary of terms, and detailed instructions for completing the template were sent to a total of 165 individuals (15 each from USF and from 10 other academic institutions). The individuals and institutions were selected from the current roster of members of the Association of Academic Surgeons, a large membership organization of academic surgeons of all academic ranks throughout the United States. First, all institutions which had at least 15 members of the Association of Academic Surgeons were identified from the current organization membership list, sorted by state and by institution. Ten institutions from this list, with at least 15 members each, were selected in order to provide geographic distribution across the United States without obvious redundancy. Within each institution selected, a random selection of 15 individuals was made, using a random number table for selection. 15 individuals from the University of South Florida, all academic surgeons, were included as an internal control, since these individuals had been exposed to the process, had attended a presentation of the methodology, and had participated in discussions regarding the use of the template. This subgroup thus had greater “training” than the remaining 150 academic surgeons.

The entire packet was submitted to each individual by direct mail. Return envelopes were included with an identification of the institution, but not the individual, on the envelope, to identify institutional origin but retain anonymity. This was carefully explained in the instructions. A number of packets were returned as non-deliverable indicating that the individual had moved to another university or had left academic surgery. Other packets were forwarded to the individual who had moved, and these were identified with the initial institution to which the packet was sent, as the individual had had documented experience at that institution in the recent past (forwarding was determined by the US Post Office and this maintained the anonymity of the respondents).

Data from the external validation were entered into an Excel spreadsheet by a member of the staff to assure that any identifying information that might have been included was removed. All data were entered in coded format. Data entry was reverified to assure accuracy.

3.3.2 Internal Validation. The author and one senior surgical colleague attended Morbidity and Mortality Conference at the University of South Florida (USF) over a fifteen month period and simultaneously scored the cases which were presented by senior residents and discussed by the surgical faculty. The first three months of reviews were used as a “training period” and the two reviewers compared their observations and discussed differences in interpretation. Following the training period, data were collected over the subsequent 12 months. Cases were included if both reviewers attended the same conference and submitted a completed template after listening to the case presentation

and discussion. 72 cases were analyzed by both reviewers. The results of the templates for these cases were compared.

3.4 Case Review Methodology and Data Collection

The scoring template, used in paper format for the validation studies, was converted to an identical electronic format and combined with the mandatory weekly complication report of the USF Department of Surgery. Each week all residents assigned with senior responsibility routinely completed a complication report electronically, in which all surgical procedures performed are listed, together with all complications and deaths. In addition, with approval of the Chairman of the Department and the Program Director of the Surgical Residency, the residents are required to complete the error report as a component of the weekly Morbidity and Mortality Report. The data relating to the error reporting template are “stripped” electronically from the submitted report and automatically transferred without patient identification into a separate electronic database. No Personal Health Information is included in the error report, assuring the protection of the patients forming the basis of the report. The identity of the reporting resident is also not included in the error data report, assuring the anonymity of the resident. This database is made available to the investigator only after any and all identifying information had been electronically removed and responses coded into digital format. The Morbidity and Mortality report provides the data to create the monthly tally of the total number of surgical procedures performed and the total number of complications reported. No patient identification or personal health information is transmitted to the investigator either by the electronic data transfer or the departmental

reports. Although a subset of the cases included in this study are discussed at the Department of Surgery weekly Morbidity and Morality Conference, information from the conference is not used in this study except in the isolated circumstance of the direct comparison of the analysis of outcomes for internal validation using the templates of two specific faculty, collected over 12 months. Both faculty members have a legitimate reason to participate in the conference, independent of this study. No patient specific information, identified information, or protected information is collected or analyzed at any point in the study.

3.5 Near-Miss Reporting

Based on the review of the literature, for every complication or adverse event there should be a multiple of near-miss events (latent errors). In order to capture and track the number and types of near-miss events, an anonymous, password protected, web-based data collection system for house staff (residents) to report near-misses was instituted. The near-miss report addressed all clinical services, not just surgery, and all participating institutions. The primary purpose of the near-miss reporting system was to provide anonymous reports back to the respective institutions to inform their quality improvement processes. The web page was integrated into the house staff duty hours documentation package used by the GME program, which was very familiar with all residents. Residents were instructed about the availability of the near-miss reporting system at orientation. The House Staff Association was advised of the near-miss reporting system on three separate occasions and asked to encourage use by the residents. Each resident received three separate e-mails advising them of the availability of the

near-miss reporting system. All program directors were advised of the availability of the near-miss reporting system both by e-mail and at the annual program directors' workshop/retreat. Unfortunately, the number of near-misses actually reported was very small. Although the information was provided to the hospital leadership, it was not used in this study. Identifying the reasons for the poor response would be informative, but it is beyond the scope of this study.

3.6 Preliminary (Formative) Data Review

After three months of Morbidity and Mortality reporting data had been collected, a preliminary analysis of the data was performed to determine the most frequent types of errors. The largest number of errors was in the error classification of "technique error." Since preventing and correcting these errors form the "substance" of a surgical residency and are known to require a long time of intensive training (the surgical residency program, accredited by the ACGME, requires sixty months of structured, formal training), "technique error" was not considered for the focused and time-limited simulation training used in this investigation. The next three most frequent causes of error were selected as the basis of a limited training program: judgment error, carelessness/inattention to detail, and incomplete understanding of the problem.

3.7 Designing the Training Program

The American College of Surgeons (ACS) provided invaluable assistance in the development of the training program in the form of participation by the investigator in

two formal committees: the Professionalism Committee and the Allied Health Professions Patient Safety Subcommittee.

The investigator served on the Professionalism Committee for two years and learned how to identify, plan, and develop video-based training tools for training surgeons in the area of “Professionalism.” During this training, the investigator worked with a group of nationally recognized surgical experts in identifying the components of Professionalism, writing scripts for vignettes for each of the identified components of Professionalism, developing the scenes, performing the taping and editing, and doing a post-development analysis of the principles and concepts which had been addressed. This experience formed the background for the “Taped Scenario” component of the training program which was developed for this study. In addition, the Professionalism Committee reviewed the scripts developed for the video component of the training.

The investigator chaired the Allied Health Professions Patient Safety Subcommittee. This subcommittee includes individuals from a variety of healthcare disciplines which all participate in the care of surgical patients: surgeons, an anesthesiologist, nurse practitioners, a physician assistant, an operating room nurse manager, surgical educators, and surgical simulation experts. This committee participated in a series of conference calls which defined and amplified a short-term goal and a long-term goal for the American College of Surgeons. With the approval of the leadership of the ACS, the short term goal was to assist in the basic design of a simulation-based training program which could serve as the prototype for the

investigator's independent work. The long-term goal was the development of a CD which would include a number of video scenarios to be used in physician training.

The Subcommittee identified four clinical sites for simulation (operating room, recovery room, emergency room, and intensive care unit). The Subcommittee also accepted using the three types of error determined in the preliminary phase of this investigation as the focus of the four vignettes (judgment, incomplete understanding of the problem, carelessness/inattention to detail). The group then met at the American College of Surgeons headquarters in Chicago and outlined potential scenarios for the four clinical sites. Equal effort was invested in developing each of the four scenarios and included identifying possible errors, defining the individuals who should be included in the vignette, and addressing how each of the errors could be accomplished in a videotaped vignette so as to be easily understood, unambiguous, and of educational value.

3.7.1 Developing the Videotaped Scenarios. An initial script, based on error in the operating room (Appendices 8, 9,10), was created by the investigator and further developed and edited by the Patient Safety Subcommittee. The script was iteratively circulated to the subcommittee by e-mail for suggestions and comment. The final script was presented to the Professionalism Committee of the ACS for input and comment. The Professionalism Committee was asked as expert individuals to anonymously “grade” the final script based on the likelihood that it would be an effective training tool on a scale of 1-10.

The finalized operating room scenario script was used to produce the videotape, utilizing resources available to the investigator through the College of Medicine, and utilizing faculty and students from the Department of Theater in the College of Visual and Performing Arts at USF. The scripts were reviewed by two senior faculty members in the Department of Theater and minor, non-substantive changes were made to enhance the dramatic quality, based on the expertise of the faculty members. All videos were taped on the same day and edited. The final videotape was presented to the Patient Safety Subcommittee for input and comment. The subcommittee voted unanimously that the videotape accurately and clearly described the three selected types of error and would be effective as a training device.

3.7.2 Developing the Mannequin-Based Simulation Training. In addition to the video, described above, the training module was designed to include an interactive, live training session using a programmed physiological mannequin (“Stan”- Appendix 11). The recovery room (Post Anesthesia Care Unit-PACU) scenario (Appendix 12) developed in the Patient Safety Subcommittee was used as the background material for the development of this simulation-based training module. To increase the validity of the simulation, a series of meetings was held which included the investigator, a professor of Industrial Organizational Psychology, Michael Brannick PhD, and two psychology doctoral students, for the purpose of planning the training session. Professor Brannick had previously published books and articles on the subject of training and is considered an expert in the field. The plan was to create a doctoral research project in psychology which would be “nested” within the training session. The collaboration in development and validation of the training session was a critically important component of this work.

For the purposes of this study, however, the actual analysis of the performance of the residents, the debriefing, and the pre-test/post-test form the basis of a separate doctoral project, are not included as a component of this study, and do not appear in this manuscript.

The USF Simulation Center agreed to assist in the development and implementation of the resident training session and approved the use of the simulation center (Center for Advanced Clinical Learning) for the actual sessions. The Center also provided both staff and expertise in the development and implementation of the simulation-based training. In a preliminary meeting with the director of the Center and her staff, the content of the simulation training session was reviewed, an extensive review of possible responses was codified, needed supplies were identified, a script was developed (Appendix 12) and the mannequin was programmed to produce appropriate physiological responses to the interventions anticipated to be requested by the trainees. The 34 available surgical residents were scheduled randomly for one of two selected days for training. Residents were scheduled to arrive at 15 minute intervals to avoid individuals meeting and discussing the session. Each resident was allocated one hour for the training. Three residents did not present for their assigned training session and were rescheduled for the subsequent session. Each resident was scheduled to participate in four fifteen minute segments: 1. the introduction, informed consent, and pre-test; 2. watching the video (12 minutes); 3. performing the simulation-based training (15 minutes); 4. debriefing by a senior surgical faculty member (15 minutes). Six weeks after the training session, the residents were asked to complete the open-ended test a second time.

3.8 Institutional Review

Consultation with the legal advisor to the University of South Florida Institutional Review Board established that the data collection process was an extension of the existing Morbidity and Mortality review process and not a new collection of research data, in that it was a refinement of an ongoing process rather than collection of new information. In addition, all data were de-identified before transfer within the server. However, it was clear that the training session for the residents involved a protected class of subjects, and that any information obtained from the training session would qualify as “generalizable research.” Accordingly, the study, including the training module and the associated evaluations, was submitted to the USF IRB and approved.

3.9 Data Analysis and Statistics

Results of surveys and templates were accumulated electronically in an Oracle database and downloaded to an Excel spreadsheet for review and analysis. Summary statistics were determined and multivariate comparisons were performed. Data are expressed as mean, median, quantiles, standard deviation, and SEM. Commercial statistical packages (MiniTab, SPSS, SAS, True Epistat) were used as appropriate to perform Fisher’s Exact tests of frequency data, Kolmogorov Smirnov tests of normality and of distribution comparison, linear regression, trend line analysis, and ANOVA. Parametric statistical tests were employed only when normality testing affirmed no significant difference from a normal distribution and examination of subsequent residuals demonstrated the reasonableness of the assumption of normality. Conditional probability

for individual analyses was determined by measuring the actual frequency of events in a subgroup as a percentage of the overall frequency of group events.

3.9.1 Reliability Analysis. Reliability is the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials. Although many studies utilize Cronbach's alpha or the kappa statistic to assess reliability, was selected. This is the sum of agreement (true positive plus true negative) divided by the total number in the subgroup at risk. They describe this as not being subject to bias and suggest a critical value greater than 0.7 to confer "reliability" on classification/taxonomy systems. In this study, reliability was assessed between two experts who "scored" a series of actual case presentations, among respondents to the set of three standard scenarios, and between aggregate data from the individual scenarios.

3.9.2 Validity Assessment. Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. External validity refers to the extent to which the results of a study are generalizable or transferable. Internal validity refers to (1) the rigor with which the study was conducted (e.g., the study's design, the care taken to conduct measurements, and decisions concerning what was and wasn't measured) and (2) the extent to which the designers of a study have taken into account alternative explanations for any causal relationships they explore.

3.9.2.1 Face Validity. Face validity is concerned with how a measure or procedure appears. Does it seem like a reasonable way to gain the information the researchers are attempting to obtain? Does it seem well designed? Does it seem as though

it will work reliably? Face validity was ascertained by demonstrating the instruments and materials to designated groups of content experts. For the template, the group of experts was the faculty of the Department of Surgery at the University of South Florida. For the video, the groups of experts were the Patient Safety Subcommittee of the American College of Surgeons, the Professionalism Taskforce of the American College of Surgeons, and the faculty of the Department of Theater at the University of South Florida. For the simulation, the groups were the Patient Safety Subcommittee of the American College of Surgeons, and the faculty and graduate students of the Division of Industrial Organizational Psychology in the Department of Psychology at the University of South Florida, and the professional simulation staff of the Office of Curriculum and Medical Education at the University of South Florida.

3.9.2.2 Construct Validity. Construct validity seeks agreement between a theoretical concept and a specific measuring device or procedure. Construct validity of the template was assessed by comparing the scores of the expert surgeons evaluating the scenarios by survey with the planned scores of the design scenarios.

3.9.2.3 Criterion Validity. Criterion related validity, also referred to as instrumental validity, is used to demonstrate the accuracy of a measure or procedure by comparing it with another measure or procedure which has been demonstrated to be valid. Criterion validity for the types of error was assessed by comparing the scores in individual components of the analysis for the individual type of error with the previously accepted classification of error in the categories of evaluation versus execution, slip, and mistake as defined previously. The types of errors identified throughout the study period

consistently related to the published characteristics of errors, that is, whether they occurred during evaluation or execution, and whether they were a slip or a mistake.

3.9.2.4. Content Validity. Content validity is based on the extent to which a measurement reflects the specific intended domain of content. The classification system used throughout this study is based on the prevalent taxonomies published over the past 25 years and is modified only to use language and concepts that are more familiar to surgeons. Errors are classified into organizational/systemic, technical/mechanical, and human factors categories, as is now widely accepted. Since neither the communication with local and national experts nor the pilot data collection suggested that there would be a large component of organizational or mechanical errors, these components were not further divided into subcategories. Anticipating that human factors errors would represent the dominant class of errors, and planning to use human factors errors in the training program for surgeons, human factors error was subdivided into 9 categories, reflecting the published literature in the field.

3.9.3 Data Classification Analysis. Responses to scoring templates were analyzed in 3 ways: overall descriptive analysis of all data; by grouping of four error types into a single category (carelessness/inattention to detail, judgment, incomplete understanding, and technical error) to estimate the previously described categories of slips and mistakes; by grouping of three error types into a single category (carelessness/inattention to detail, judgment, incomplete understanding) to estimate the three categories that were specifically designed into the training instrument. In other words, data were structured and analyzed to facilitate the comparison of data after training to baseline data, and to

specifically address 1. overall change (all error categories), 2. slips and mistakes, represented by the four categories listed, 3. areas of specific training, represented by the three categories.

Regressions were performed to analyze whether number of procedures, number or percent complications, percent complications associated with error, or percent complications associated with index errors changed during the 12 month baseline period. Normality of the data was affirmed by analysis of residuals. Regressions and ANOVA were performed using MiniTab or SAS. Individual distributions were assessed for normality by visually examining graphs of residuals and by Kolmogorov-Smirnov testing. In addition to traditional statistical hypothesis testing, pre- and post-training data were compared by ANOVA of linear regression. For purposes of hypothesis testing, the null hypothesis was rejected if $p < 0.05$ unless otherwise specified.

4. Results

The original prototype of the scoring template of medical errors was created based on a thorough review of the published literature on industrial error, medical error, and surgical complications. The prototype was refined utilizing the input of internal academic surgical colleagues and tested at the surgical Morbidity and Mortality Conference. The list of medical errors was sent to eight external academic surgical experts qualified by having published peer reviewed articles on the subject of patient safety. Four experts responded to the request (50%) after two mailings. All experts agreed with the classification system.

The case scenarios were reviewed with members of the surgical faculty, to verify that they were realistic and that the medical error appeared to be clear and scorable using the template. Adjustments were made to facilitate comprehension and ease of use. The template was also reviewed by two nationally recognized experts (S.M., College of Nursing; J.K., College of Education) in measurement and evaluation to verify that the instrument itself was understandable, consistent, and likely to be usable for external validation.

One hundred sixty-five scenarios and templates were sent out as described to selected members of the Association of Academic Surgeons. A second mailing to each individual, thanking him/her if he had completed the survey and reminding him/her if not

yet completed, was sent 6 weeks after the first mailing.. After two mailings, 48 completed evaluations were returned completed (29% response rate).

4.1 External Template Validation

Forty-eight completed evaluations were available for analysis. The average age of the respondents was 15.9 +/- 9.2 years post surgical residency. All respondents had current faculty appointments in a major university and were actively involved in academic surgery (Table 2).

Table 2. Years Since Completing Residency

Years Since Residency	N
In Residency	3
0-5	2
6-10	9
11-15	10
16-20	7
21-25	8
26-30	4
31-35	4
36-40	0

As designed, each of the three scenarios was intended to demonstrate a specific type of error: Scenario 1, technical error (31) and judgment (28); Scenario 2, carelessness/inattention to detail (25); and Scenario 3, judgment (28). In addition, each describes a different severity of injury and contains a differing level of complexity. Scenario 1 is straightforward with minimal impact on the patient. Scenario 2 is complicated (multiple components but a prudent individual will identify them correctly) and would result temporary harm to the patient. Scenario 3 is complex (multiple components which could be analyzed in several ways) and would lead to serious harm to the patient. The Outcome Scores for the three scenarios are summarized in Table 3.

Table 3. Outcome Score for the Three Scenarios
Patient outcome reported by surgeon experts for three clinical scenarios.

		Scenario 1		Scenario 2		Scenario 3	
Outcome Definition	Outcome ID	#	%	#	%	#	%
No injury or delay	1	15	34.1	1	0.0	0	0.0
No injury but increased length of Stay (LOS)	2	23	52.3	8	16.0	0	0.0
Injury with temporary disability	3	6	13.6	39	78.0	19	44.2
Injury with permanent disability	4	0	0.0	2	4.0	24	55.8
Death	5	0	0.0	0	0.0	0	0.0

If the sum of Outcome Scores 3+4+5 is considered to represent serious adverse outcomes (National Association of Insurance Adjustors criteria), then 14% of responders scored Scenario 1 as having a serious adverse outcome, 82% scored Scenario 2 as having a serious adverse outcome, and 100% scored Scenario 3 as leading to a serious adverse outcome. This confirms the ability of the scoring system to distinguish the severity of the error. This is particularly noteworthy since none of the patients died in the scenarios, eliminating “death” as an outcome score. The weighted score for Scenario 1 is 1.8, Scenario 2, 2.9, and Scenario 3, 3.6. This estimate of the impact of the events in the scenario on patient outcome seems to be appropriate in that the patient in Scenario 1 might have slight prolongation in length of stay (outcome score less than 2); the patient in Scenario 2 would have some disability which might be temporary (outcome score approximately 3); and the patient in Scenario 3 would be expected to have a very prolonged hospitalization with multiple complications, probably requiring subsequent surgical procedures (outcome score near 4). The scoring template appears to capture the impact of the error on patient outcome realistically. Listed in Table 4 is the summary information on error occurrence for the three scenarios.

Table 4. Agreement in Error Classification in the Scenarios

	Weighted							Wt. %
	Score	% Error	% Eval	%Exec	Both	Slip	Mistake	Impact
Scenario 1	1.8	77%	8.1%	92.0%	0.0%	93.5%	6.5%	65.8%
Scenario 2	2.9	98%	2.2%	93.3%	4.4%	41.2%	58.8%	83.9%
Scenario 3	3.6	79%	50.0%	44.1%	5.9%	42.9%	57.1%	58.0%

By design, each of the scenarios represented an error. In the experience of the author, many surgeons interpret a surgical consequence that is difficult to avoid to be “unavoidable” and thus perhaps not due to error. An injury to the intestine during an operation is a case in point. While it is an error, in the sense that it is an unintended consequence which could have been avoided, in scenario 1 approximately 25% of surgeons did not identify the intestinal injury as an error. Similarly, 21% of surgeons did not identify an error in Scenario 3, while fully 98% identified an error in Scenario 2. Scenario 1 and scenario 2 represent strictly technical “failures” and as would be expected, greater than 90% of responders identified an error occurring during execution. Scenario 3 was identified by half of surgeons as an error during evaluation and half as an error during execution. This very complex case contains elements of both types of error. The template fails to be able to resolve this dilemma in the simple binary choice of evaluation or execution. This complexity is more readily apparent in the more detail analysis of specific error types presented and discussed later. In Scenario 1, the surgeons uniformly identified the error as a slip (the correct action done incorrectly). In Scenarios

2 and 3 there was an almost even assignment to slip and mistake, even when the error was clearly identified as occurring during execution as in Scenario 2.

The more detailed analysis of the types of error provided by the broader error classification scheme in the template is demonstrated in Table 5 and graphically depicted in Figure 1.

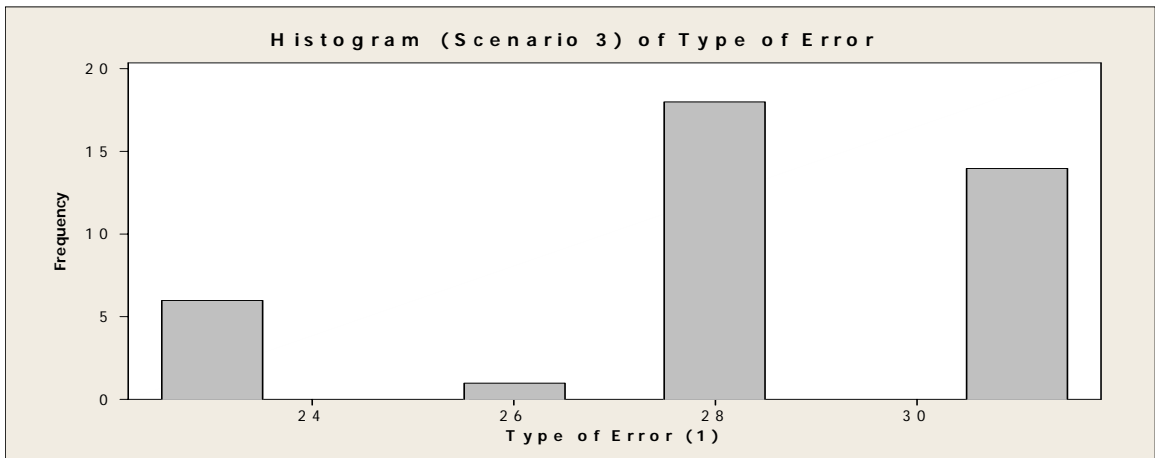
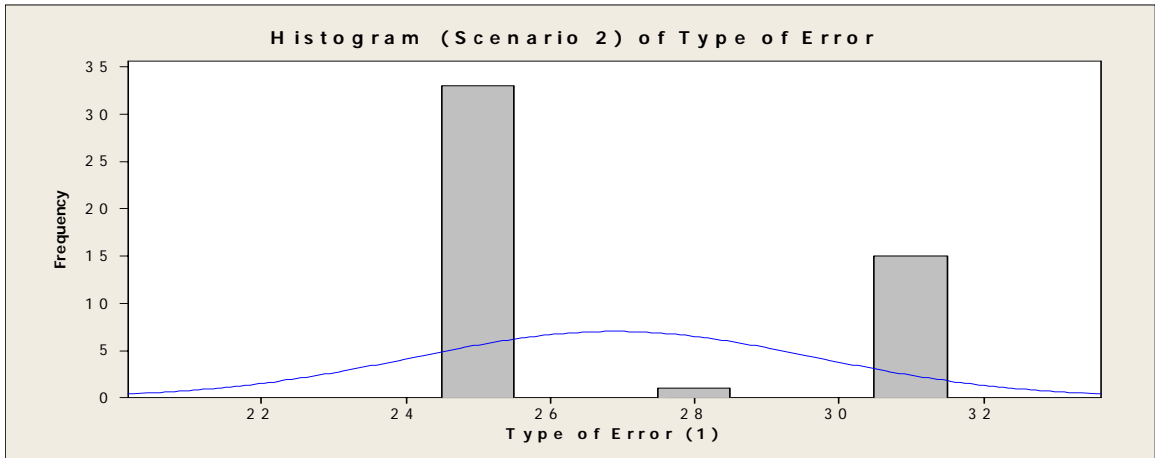
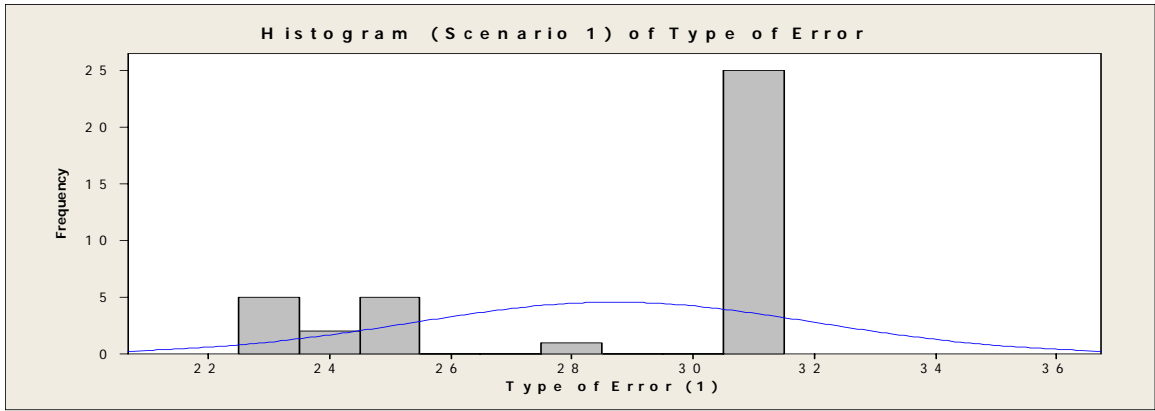


Figure 1. Error Types Identified in Scenarios

The bars indicate the frequency that specific errors were identified for each of the three validation scenarios.

Table 5. Characterization of Error Types in Validation Scenarios

For each scenario, the percentage of expert responses identifying the primary source of error.

Error Types	Error ID	Scenario 1	Scenario 2	Scenario 3
Equipment/Mechanical	21	0%	0%	0%
Organizational error	22	0%	0%	0%
Incomplete understanding	23	13%	0%	15%
Failure to use protocol	24	5%	0%	0%
Carelessness/inattention to detail	25	13%	67%	0%
Error in diagnosis	26	0%	0%	3%
Communication error	27	0%	0%	0%
Judgment error	28	3%	2%	46%
Delay error	29	0%	0%	0%
Error of omission	30	0%	0%	0%
Technique error	31	66%	31%	36%

This table only includes the *primary error* identified for each scenario. The numbers in bold identify the major errors identified by the surgeons for each scenario. For Scenario 1, most responders felt that this was a technical problem, although approximately 25% identified the error as either incomplete understanding or carelessness/inattention to detail. In Scenario 2, 67% identified the error as carelessness/inattention to detail, although almost one third saw this as an error in technique. Similarly, in Scenario 3, almost 50% recognized a judgment error, yet a third felt that this was a technique error. This information for Scenario 2 and Scenario 3

provides a better description of the error than the simple classification into slip or mistake, which would appear to be “confounding” the combination of more than one error.

Table 6. Characterization of All Errors Listed in Validation Scenarios
For each scenario, the distribution of all errors reported by experts.

Error Description	Error ID	Scenario 1	Scenario 2	Scenario 3
Equipment/Mechanical	21	0%	2%	0%
Organizational error	22	0%	0%	0%
Incomplete understanding	23	10%	6%	18%
Failure to use protocol	24	7%	5%	2%
Carelessness/inattention to detail	25	27%	39%	8%
Error in diagnosis	26	0%	1%	16%
Communication error	27	0%	1%	0%
Judgment error	28	11%	10%	30%
Delay error	29	0%	3%	0%
Error of omission	30	0%	3%	0%
Technique error	31	45%	30%	26%

Table 6 includes *all errors* identified by the surgeons for each of the three scenarios. In this analysis, there is no attempt to reconcile the order of the errors (which will be addressed below). Although technique is identified by 46% in scenario 1, this is less than when only the primary error is considered, thus allowing a greater appreciation of the cognitive or planning error that was designed into the scenario. Scenario 2 is actually made more confusing by considering all listed errors without weighting, decreasing the “carelessness/inattention to detail” from 66% to 39% and introducing a variety of additional errors (equipment/mechanical, incomplete understanding, failure to use protocol, error in diagnosis, communication error, judgment error, delay error, and

error omission). This would not appear to be helpful. Scenario 3, an even more complicated clinical situation, is similarly more diffuse when all errors are considered. When the data are combined into a tabular form and apply weighting for the location of the error types in the individual surgeon's list (Table 7), the major error is more clearly accentuated and yet the associated error types are also clear.

Table 7. Scenario Data Analysis Weighted by Position in List

Case 1	List	E21	E22	E23	E24	E25	E26	E27	E28	E29	E30	E31
Weight = 1	5 th	0	0	0	0	0	0	0	0	0	0	0
Weight = 2	4 th	0	0	0	0	0	0	0	0	0	0	1
Weight = 3	3 rd	0	0	0	1	4	0	0	4	0	0	2
Weight = 4	2 nd	0	0	2	2	12	0	0	4	0	0	3
Weight = 5	1 st	0	0	4	2	5	0	0	0	0	0	27
Sum		0	0	6	5	21	0	0	8	0	0	33
Weighted sum		0	0	28	21	85	0	0	28	0	0	155
Weighted ave		0	0	1.87	1.40	5.67	0.00	0.00	1.87	0.00	0.00	10.3

Case 2		E21	E22	E23	E24	E25	E26	E27	E28	E29	E30	E31
Weight = 1	5 th	0	0	2	1	0	0	0	0	0	0	0
Weight = 2	4 th	1	0	3	0	0	0	0	2	0	1	1
Weight = 3	3 rd	0	0	1	2	2	1	1	6	2	2	3
Weight = 4	2 nd	0	0	1	3	12	0	0	3	2	0	13
Weight = 5	1 st	0	0	0	1	28	1	0	1	0	0	16
Sum		1	0	7	7	42	2	1	12	4	3	33
Weighted sum		2	0	15	24	194	8	3	39	14	8	143
Weighted ave		0.13	0.00	1.00	1.60	12.9	0.53	0.20	2.60	0.93	0.53	9.53

Case 3		E21	E22	E23	E24	E25	E26	E27	E28	E29	E30	E31
Weight = 1	5 th	0	0	0	0	1	0	0	0	0	0	0
Weight = 2	4 th	0	0	0	1	0	1	0	0	0	0	2
Weight = 3	3 rd	0	0	3	0	4	2	0	5	0	0	3
Weight = 4	2 nd	0	0	7	1	3	11	0	2	0	0	5
Weight = 5	1 st	0	0	5	0	0	2	0	19	0	0	14
Sum		0	0	15	2	8	16	0	26	0	0	24
Weighted sum		0	0	62	6	25	62	0	118	0	0	103
Weighted ave		0	0	4.1	0.4	1.7	4.1	0	7.9	0	0	6.9

Scenario 1 is clearly identified as technical error, yet carelessness /inattention to detail (and to a lesser extent incomplete understanding of the problem and judgment error) are also recognized. Scenario 2 is recognized as principally carelessness/inattention to detail and a technical error. Scenario 3 is a judgment error and a technical error, but also identified is the fact that there is perhaps an error in diagnosis. Although the total number of respondents is relatively small (N=45), the analysis of the data would suggest that the classification system can provide both a valid and a rich understanding of the errors.

Twelve completed scenarios were submitted by faculty and residents at the investigator's institution (University of South Florida) and were also analyzed independently to determine if reliability and validity within the study institution was greater than of the sample as a whole. Since the individuals within the investigator's institution had the greatest familiarity with the instrument, it might be anticipated that this group would more consistently use the scoring template to evaluate error (reliability and validity). The average number of years since residency for this group (8.1 +/- 2.0 years) is significantly less than for the other respondents (15.9 +/- 1.3 years) or the total group (18.3 +/- 1.4 years). $p < 0.05$ for both comparisons. This is in part due to the inclusion of 3 residents in this group (there were no residents from any other institution), yet the difference is still significant ($P = 0.05$) if the residents are excluded. The analysis of this subgroup, which is perhaps more indicative of the group that participated in the actual patient error reporting (to be reported below), is considerably more homogeneous than the larger group as a whole. For Scenario 1, 82% of respondents identified an error, all noted it occurred during execution, as a slip. For Scenario 2, 100% identified the action

as an error during execution, and a slip. For Scenario 3, 73% identified an error; half placed it in evaluation (and all noted a mistake) and half in execution (and all noted a slip). The outcome scores were, respectively, 1.7, 3.1, 3.2 with Scenario 2 slightly but not significantly higher than the total group, and Scenario 3 slightly lower than the total group but not significantly.

4.2 Internal Comparison

Seventy-two patient presentations were scored by each of two senior surgical faculty members over a 12 month period. Cases were selected for presentation in the conference by the surgical residency program director independently, without regard to the study. Cases were only scored when both faculty were present at the same Morbidity and Mortality Conference to evaluate the same case presentation. This methodology, while not random, is unbiased.

Both evaluators agreed on the presence or absence of error in 59 of the 72 cases (reliability = 0.83). By Fisher's Exact Test, the probability was 0.41 (NS). The agreement between the two evaluators and the breakdown of outcome scores are indicated below in Tables 8 and 9.

Table 8. Agreement Between the Two Evaluators

	Evaluator 2 No	Evaluator 2 Yes
Evaluator 1 No	24	9
Evaluator 1 Yes	3	36
Reliability = 0.83		
Exact-P=.4097		

Table 9. Internal Evaluator Outcome Scores

Outcome was scored on a 1-5 scale.

SCORE	Description	Evaluator 1		Evaluator 2	
1	No adverse effect on outcome	2	0.03	7	0.10
2	No injury but increased length of stay	13	0.18	14	0.19
3	Injury with temporary disability	33	0.46	27	0.38
4	Injury with permanent disability	5	0.07	5	0.07
5	Death	19	0.26	19	0.26
		Weighted score	3.36	Weighted score	3.21

Graphically, the outcome scores of the two evaluators demonstrate a high degree of agreement (Figure 2).

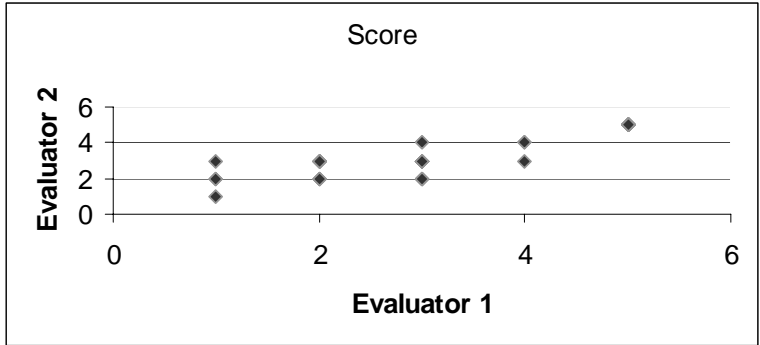


Figure 2. Comparison of Error Scores (1-5) Between Evaluators 1 and 2

The outcome scores of the two evaluators were compared by linear regression and ANOVA. There was a very high agreement between the evaluators:

$$\text{Score \#2} = 0.99970 (\text{Score \#1}) - 0.1518 \quad (p < 0.0001)$$

Correlation analysis demonstrated $r = .88$ and ANOVA demonstrated an adjusted R-squared = 0.77. Analysis of residuals (Figure 3) showed no evidence of non-normality or unequal variance. KS analysis for Evaluator 1 demonstrated mean outcome score = 3.36, SD 1.14, median 3, KS=0.069 ($p > .15$); for Evaluator 2 mean outcome score=3.208, SD=1.30, median 3, KS=0.055 ($p > .15$). Using the KS test to compare the two distributions of scores $p > .99$. Scores were also compared by Fleiss Kappa statistic and no difference was detected in any of the five outcome score levels. There does not appear to be a significant difference between the scoring processes of the two evaluators.

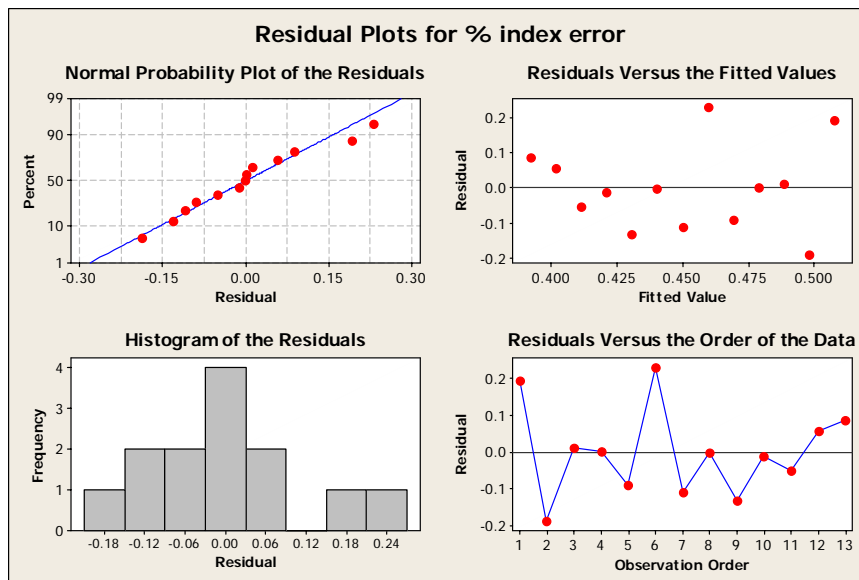
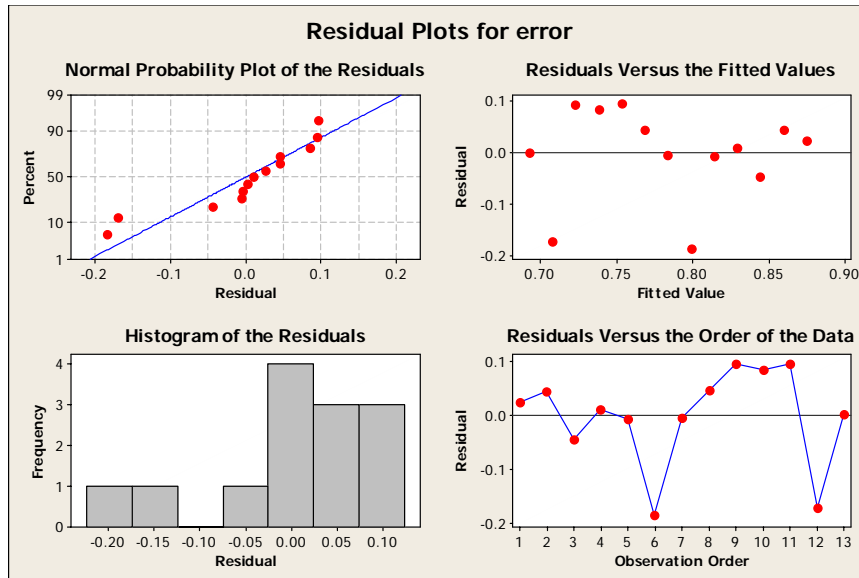


Figure 3. Residual Analysis of Error and Index Error

Residuals were calculated for the trend analysis of total error and index error over time for the entire study period.

The assessments of the two evaluators were compared as to whether they believed the error occurred during evaluation or execution. Previous studies have separated errors

that occurred during evaluation (planning) from those occurring during execution. This generally corresponds to the distinction between mistakes (planning) and slips (execution). The evaluators identified very few cases where error had occurred during evaluation (Evaluator 1 in 6, Evaluator 2 in 4) making the analysis of errors during evaluation “trivial.” In those cases where the two evaluators agreed on the presence of error (N=36), both evaluators agreed that the error occurred during execution in 27/28 opportunities but only agreed in 50% that the error occurred during evaluation in the remaining cases. This could have been confounded by the use of the term evaluation instead of planning, as the missing cases leading to lack of agreement were usually not scored in this category. Both evaluators therefore consistently agreed on the presence of errors of execution and likely agreed similarly on the presence of errors of evaluation/ planning.

The assessment of the two evaluators as to whether the error was a slip or a mistake was evaluated in the 36 cases where the evaluators agreed that an error had occurred. The evaluators demonstrated a significant agreement regarding the occurrence of a slip versus a mistake. In 31 cases where the error had occurred during execution, Evaluator 1 identified a slip in 25 and Evaluator 2 in 26. The evaluators identified very few errors occurring during evaluation. However, of the 6 errors during evaluation which were identified by Evaluator 1, one was classed as a mistake and not classified in the other 5. Evaluator 2 identified a mistake in 3 and a slip in 1, in the 4 evaluation errors identified. This is not entirely conclusive because in 5 cases neither a slip nor a mistake was coded. These data indicate a high reliability in identifying the presence of a mistake

in errors of evaluation and of a slip in errors of execution, demonstrating internal consistency in the evaluation template scoring.

The primary error type for the two evaluators is categorized in Table 10.

Table 10. Distribution of Error Types in Inter-rater Comparison

Error ID	Eval 1	Eval 2	<i>Error Definition</i>
	<i>Frequency</i>	<i>Frequency</i>	
21	1	1	equipment/mechanical failure
22	1	0	organizational error
23	1	9	incomplete understanding
24	3	2	failure to use protocol
25	11	8	carelessness/carelessness/inattention to detail
26	1	1	error in diagnosis
27	1	3	communication error
28	2	2	judgment error
29	1	1	delay error
30	0	0	error of omission
31	17	18	technique error

Graphically, Evaluator 2 was more likely to identify that there was incomplete understanding of the problem, whereas Evaluator 1 was more likely to identify carelessness/carelessness/inattention to detail.

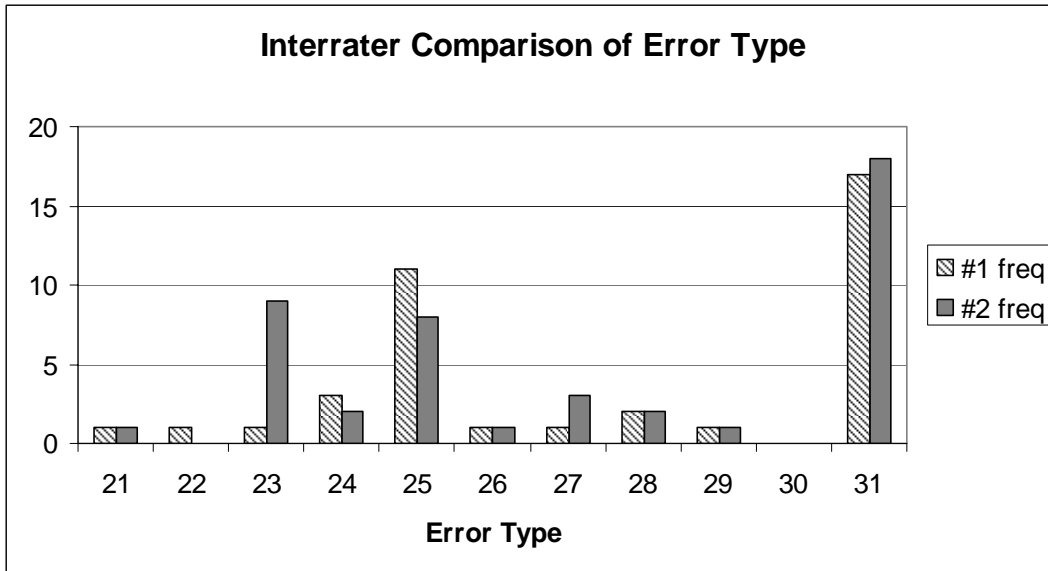


Figure 4. Graphical Comparison of Errors Identified by Internal Raters

The evaluators agreed on the exact primary error type in 64% or 23/36 cases (Figure 4). The agreement was very high when all error types listed for a case were compared between the two investigators, specifically identifying the likelihood that the first error type was one of the four types which corresponded with the previously described categories of slip and mistake (Error ID 23,25,28,31). The evaluators agreed in 92% or 33/36 cases in which an error had occurred that the primary error was one of these four. The specific training was designed to address three of these four, incomplete understanding, carelessness/carelessness/inattention to detail, and judgment, and was not specifically designed to address error in technique. When the three error types (23,25,28) were compared between the two evaluators, there was agreement in 72%.

4.3 Video Evaluation

4.3.1 Video Script Review. Thirteen nationally recognized medical educators, all members of the Professionalism Task Force of the American College of Surgeons, reviewed the script and responded to the question “How likely is this process to be an effective means of training residents in the area of medical error?” They were instructed to write down (anonymous) a score from 1-10. The average score was 8.5 ($p < 0.05$) with a standard error of the mean of 0.26 (95% confidence interval 7.9-9.0)

4.3.2 Video Review. The completed video was demonstrated to the Committee on Allied Health Professionals of the American College of Surgeons at the Annual Clinical Congress of the American College of Surgeons. Ten nationally respected academic surgeons and two non-physician surgical educators participated. The members were asked to provide feedback and criticism. The committee voted unanimously to endorse the video as a valid training instrument, with a high likelihood of being effective in training surgical residents in the area of medical error.

4.4 Analysis of Baseline Error Data

For 3 months, morbidity and mortality templates were reviewed by the investigator to assure that the “system” was functioning, the data were being stored in a complete and retrievable manner, and that the residents were completing the assessments appropriately.

After the initial 3 months of preliminary data collection, data were collected automatically using the electronic template. A preliminary analysis of the data to determine the most frequent types of error was conducted at the 6-month time point.

The monthly error incidence is shown in Figure 5. The most common type of error in the preliminary analysis was technical, followed by judgment error, carelessness/inattention to detail, and incomplete understanding. This preliminary analysis was used to determine the material to be developed in the simulation training. The analysis was repeated at the conclusion of twelve months and demonstrated the same distribution of errors.

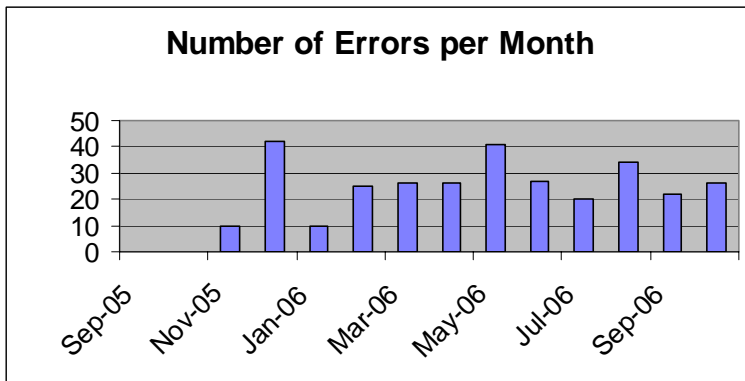


Figure 5 . Graph of Number of Errors Per Month During Baseline

Over the subsequent 12 months, all reported complications were scored using the electronic template. During the 12 month baseline analysis period, 9830 surgical cases were performed and available for analysis. A total of 332 complications were reported (3.4% reported complication rate) and an error was reported to be associated in 79% of the reported complications (Table 11). These values are consistent with data reported in the surgical literature.

Table 11. Monthly Procedures, Complications, and Errors During Baseline Period

Month	Procedures #	Complications #	Errors #	Error % %
Nov-05	798	10	9	90.0
Dec-05	767	42	38	90.5
Jan-06	927	10	8	80.0
Feb-06	761	25	21	84.0
Mar-06	811	26	21	80.8
Apr-06	748	26	16	61.5
May-06	929	41	32	78.0
Jun-06	709	27	22	81.5
Jul-06	822	20	17	85.0
Aug-06	902	34	28	82.4
Sep-06	857	22	18	81.8
Oct-06	799	26	14	53.8
<i>Total – 12 month</i>	<i>9830</i>	<i>309</i>	<i>244</i>	<i>79.0</i>
% Complications		3.1%		
Error Rate per Complication		79.0%		

The distribution of type of error is shown in Figure 6. The error was due to a slip in 58% and a mistake in 20%. The most common type of error was “technical error,” reported in 63.5% of errors. Judgment errors were reported in 29.6%, carelessness/inattention to detail in 29.3%, and incomplete understanding in 22.7%.

Since more than one type of error could be reported, the total is greater than 100%. This breakdown of errors was used for the development of the training program.

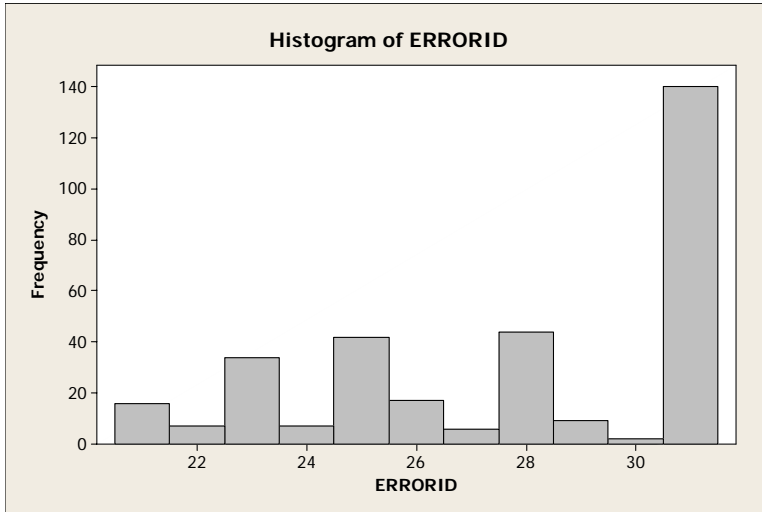


Figure 6. Frequency Distribution of Error Types During Baseline Period

The average outcome score was 2.76, and was distributed thus: no adverse effect in 3.6%, prolongation of hospitalization but no patient injury in 34.4%, definite but non-permanent disability in 25.1%, permanent disability in 8.4%, and death in 16.0%. Using a definition for serious error as the sum of 3+4+5 outcome scores yields a serious error rate in the baseline period of 49.5%. The outcome score was not different (Mann-Whitney test) in the presence or absence of an error. Patient outcome by month was analyzed by linear regression, and there was no significant change in the number of complications per month or the percent complications per month.

$$\text{Number} = 22.5 \text{ complications} + 0.341/\text{month} \text{ (p=0.65)}$$

The percentage of complications associated with a reported error decreased over the twelve month period.

Percent Error = 92% -1.5%/month; p= 0.051

The percentage of cases for which an index error was reported did not change significantly over the period.

% index error= 54% - 0.96%/month; p=.33

Table 12 demonstrates the conditional probabilities that a specific error type was reported given that the error was reported as having occurred during evaluation or execution, and given that it was reported as a slip or as a mistake. Error Types 23, 26, and 28 (incomplete understanding, error in diagnosis, judgment error) were most likely to have been reported as occurring during evaluation. Error types 25 and 31 (carelessness/inattention to detail, technique error) were most likely to have been reported as occurring during execution.

Table 12. Analysis of Characteristics of Individual Error Types

Conditional Probability of Error Type given Error Characteristics (Evaluation versus Execution and Slip versus Mistake).

Error Type		E21	E22	E23	E24	E25	E26	E27	E28	E29	E30	E31
Eval & Exec N=257	Tot Err%	0.05	0.02	0.07	0.02	0.16	0.05	0.02	0.12	0.02	0.00	0.46
	Both	0.00	0.11	0.11	0.00	0.22	0.22	0.00	0.00	0.00	0.00	0.33
	Slip	0.05	0.00	0.04	0.02	0.18	0.03	0.01	0.07	0.01	0.00	0.58
	Mistake	0.03	0.03	0.08	0.03	0.08	0.15	0.03	0.40	0.00	0.03	0.18
	Neither	0.08	0.04	0.15	0.04	0.13	0.04	0.04	0.10	0.04	0.00	0.35
Eval only N=57	Tot Err%	0.02	0.04	0.26	0.05	0.07	0.23	0.04	0.23	0.02		0.05
	Both	0.00	0.00	0.33	0.00	0.00	0.67	0.00	0.00	0.00		0.00
	Slip	0.00	0.00	0.31	0.13	0.06	0.19	0.06	0.19	0.00		0.06
	Mistake	0.00	0.05	0.15	0.00	0.05	0.30	0.05	0.35	0.00		0.05
	Neither	0.06	0.06	0.33	0.06	0.11	0.11	0.00	0.17	0.06		0.06
Exec only N=190	Tot Err%	0.06	0.01	0.02	0.02	0.19	0.01	0.01	0.09	0.02	0.01	0.57
	Both	0.00	0.17	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.50
	Slip	0.06	0.00	0.01	0.01	0.20	0.01	0.00	0.06	0.01	0.00	0.64
	Mistake	0.05	0.00	0.00	0.05	0.11	0.00	0.00	0.42	0.00	0.05	0.32
	Neither	0.10	0.03	0.07	0.03	0.17	0.00	0.03	0.07	0.03	0.00	0.45
Neither N=10	Tot Err%							0.20	0.10			0.70
	Both											
	Slip							0.25	0.00			0.75
	Mistake							0.00	1.00			0.00
	Neither							0.20	0.00			0.80

4.5 Development and Validation of Training Program

Initial data evaluation suggested that, after errors in technique, the most common types of errors reported were errors of incomplete understanding of the problem, judgment, and carelessness/inattention to detail. In order to maximize the likelihood of identifying an effect of training in altering error, these three specific errors were selected as the focus of the training program.

The American College of Surgeons agreed to create a subcommittee, The Patient Safety Subcommittee, of the Committee on Allied Health Professionals. This subcommittee met by conference call on three separate occasions of an hour each, and conference summaries were submitted to participants to inform the next conference call. The Subcommittee then met physically at the American College of Surgeons headquarters in Chicago for two days to “brainstorm” the development of the training program. The author chaired both the conference calls and the in-person meeting. An initial draft of the ideas generated was circulated to the members for agreement. Four potential scenarios were developed in the brainstorming process, which included the 3 index errors (incomplete understanding, judgment, carelessness/inattention to detail), and which would be enacted respectively in the operating room, recovery room, emergency room, and surgical intensive care unit.

Two of these, the operating room and the recovery room, were selected for use in the simulation-based training session. The operating room scenario was selected to be scripted, acted, and videotaped. The recovery room scenario was selected to be used in the mannequin based, hands on training session. The script of the operating room

scenario was sent to the members of the Patient Safety Committee for editing and comments. After two iterations, the script was reviewed by a senior faculty member in the Department of Psychology and by a senior faculty member in the Department of Theater.

The script was distributed at a face to face meeting of the American College of Surgeons Professionalism Task Force. This group was selected because it had already created and finalized 14 videotaped scenarios for the purposes of training surgeons and surgical residents. The Committee (14 members) was composed of medical experts in the fields of Surgery, Orthopedic Surgery, Internal Medicine, and Medical Education. Each member was given a copy of the draft script and was asked to comment and to vote on a scale of 1-10 on how likely the script would be effective as a training tool. The committee score for the expected efficacy of the video was 8.5 +/- 0.26SEM, with a minimum of 7, maximum of 10, and median of 8.

Preparing the videotape. The final script was given to two senior faculty members in the Department of Theater at the University of South Florida. They recruited an additional faculty member and six senior undergraduate students, who volunteered to act in the scenario. The Tampa VA hospital approved using an empty operating room on a Saturday afternoon. The Media Center of the University of South Florida College of Medicine agreed to loan a professional, digital videocamera. A photographer volunteered to operate the camera at the video shoot. Cue cards were prepared for the various components of the scenario. The scenario consisted of four discrete components (Appendix 8), an introduction, an operating room scene which included the three index

errors, a roundtable discussion of a moderator and three surgeon actors, and a repeat of the operating room scene with the errors corrected. The Media Center at USF performed editing of the final video. The videotape was demonstrated to the Committee on Allied Health Professionals of the American College of Surgeons at the Annual Clinical Congress. The committee voted unanimously that the video was representative of the three index errors and that it was highly likely to be effective as a training tool.

The Hybrid Simulation Training Session. The Director of the Center for Advanced Clinical Learning at the University of South Florida College of Medicine agreed to provide staff assistance and the use of the patient simulation center for the training sessions. The facilities included an examination room prepared as a recovery room, a trained nurse practitioner, remote cameras with a camera/audio operator, two additional nurses, and two senior surgeons. Thirty-four residents were each scheduled for one hour of training. A senior faculty member and two doctoral students from the Department of Psychology observed the training sessions and administered a survey examination to the residents regarding the training session. Each resident participated in four discrete phases: introduction/consent/pre-test; watching the 12 minute video; performing the mannequin-based recovery room scenario; and a one-on-one debriefing session with one of the two senior surgeons who had remotely observed the training session. Each participating resident repeated the survey test six weeks after the training session.

4.6 Post-Training Data Analysis

Post-training data were collected over a 6-month period in an identical manner as in the pre-training period. Data are depicted in 6-month blocks to facilitate comparison (Table 13). The first 2 periods are also combined as the pre-training period. Comparison between the 6 months immediately preceding training to the 6 months following training was used to have more equal population sizes, and the comparison was made using Fisher's Exact Test.

Data were also analyzed by regression analysis over the entire 18 month time period (based on the null hypothesis that there was no effect of the training). There was no difference between number of patients treated per month before and after training, either by Fisher's Exact test or by regression ($P=.35$).

Data are presented by month for the entire 18 month study period in Table 14. The table demonstrates procedures, complications, % complications, error, errors as percent of complications, errors as percent of procedures, primary index errors, any index errors listed, primary index as percent of procedures, any index as percent of procedures, primary index as percent of complications, and any index as percent of complications.

Table 13. Analysis of Error Data by 6-Month Blocks

Results for number and type of errors for each 6-month period. The first two periods comprise the base year. The third block is after training. The total period includes all 18 months. (P= Procedures; C=Comps=Complications; E=Errors).

	Procedures #/mo.	Comps #/mo	Comps %	Errors #/mo	E/C %	E/P %	Primary Index #/mo	Total Index#/mo	Index/P %	Total Index/P %	Primary Index/C %	Any Index/C %
First 6 mo.	802	26.0	3.2%	19.0	74.2	2.4	6.2	10.5	0.8	1.3	23.9	40.8
Second 6 mo.	836	26.2	3.1%	18.0	69.3	2.2	8.0	12.0	1.0	1.4	32.2	46.5
First Year	819	26.1	3.2%	18.5	71.7	2.3	7.1	11.3	0.9	1.4	28.1	43.6
Third 6 mo.	847	18.0	2.1%	13.2	72.3	1.6	5.0	7.7	0.6	0.9	27.4	41.5
Total Period	829	23.4	2.8%	16.7	71.9	2.0	6.4	10.1	0.8	1.2	27.9	42.9

Table 14. Analysis of Morbidity and Mortality Data by Month Over 18-Month Study Period

P=Procedures; C=Comps=Complications; E=Errors.

Month	Procedures #/mo	Comps #/mo	Comps %	Errors #/mo	E/C %	E/P %	Primary Index #/mo	Total Index #/mo	Primary Index/P %	Total Index/P %	Primary Index/C %	Total Index/C %
Nov-05	798	42	5.3	29	69.0	3.6	8	20	1.0	2.5	19.0	47.6
Dec-05	767	10	1.3	8	80.0	1.0	2	5	0.3	0.7	20.0	50.0
Jan-06	927	25	2.7	20	80.0	2.2	7	10	0.8	1.1	28.0	40.0
Feb-06	761	26	3.4	21	80.8	2.8	8	9	1.1	1.2	30.8	34.6
Mar-06	811	27	3.3	22	81.5	2.7	3	5	0.4	0.6	11.1	18.5
Apr-06	748	26	3.5	14	53.8	1.9	9	14	1.2	1.9	34.6	53.8
May-06	929	27	2.9	22	81.5	2.4	13	17	1.4	1.8	48.1	63.0
Jun-06	709	20	2.8	15	75.0	2.1	10	11	1.4	1.6	50.0	55.0
Jul-06	822	32	3.9	21	65.6	2.6	4	12	0.5	1.5	12.5	37.5
Aug-06	902	24	2.7	19	79.2	2.1	10	12	1.1	1.3	41.7	50.0
Sep-06	857	26	3.0	14	53.8	1.6	6	8	0.7	0.9	23.1	30.8
Oct-06	799	28	3.5	17	60.7	2.1	5	12	0.6	1.5	17.9	42.9
Nov-06	935	22	2.4	19	86.4	2.0	7	13	0.7	1.4	31.8	59.1
Dec-06	683	16	2.3	12	75.0	1.8	3	5	0.4	0.7	18.8	31.3
Jan-07	851	18	2.1	13	72.2	1.5	5	7	0.6	0.8	27.8	38.9
Feb-07	826	20	2.4	15	75.0	1.8	6	9	0.7	1.1	30.0	45.0
Mar-07	900	16	1.8	11	68.8	1.2	3	4	0.3	0.4	18.8	25.0
Apr-07	888	16	1.8	9	56.3	1.0	6	8	0.7	0.9	37.5	50.0
SUM	14913	421		301								
AVERAGE			2.8%		71.5%	2.0%	6.39	10	0.8%	1.2%	27.9%	42.9%

There was no difference in the number of errors, the number of primary index errors, or total index errors by Fisher's Exact Test between the 6 month pre and post-training periods. However, there was a significant regression with a decrease in the number and percent of complications and the number and percent of errors, which spanned the entire 18 month period (Table 15).

Visual evaluation of the graphical data suggests that the decreases began before the training and continued in a linear fashion after the training. Specifically, there was no abrupt discontinuity in the data following the training. Of more interest is the observation that there was no difference in the percent of Index Errors following training either by Fisher's Exact Test or by regression. In fact, it would appear that there was either no decrease or an actual increase in the index errors following training. Data are graphically depicted in Figure 7, Surgical Procedures by Month and Monthly Complications and Errors; Figure 8, Errors as Percent of Procedures and Percent of Complications; Figure 9, Primary and Total Index Error by Procedure; and Figure 10, Primary and Total Index Error by Complication.

Table 15. Regression Analysis of Principle Components Over Time

Regressions were performed over the 18-month study period. The number of procedures remained constant while the complication rate and error rate fell. The error rate per complication was stable, and the rate of index errors did not change over the study period. (P=Procedures; C=Complications; E=Errors; p = Probability).

Value per Month	Mean	SEM	SD	Min	Median	Max	Skewness	Kurtosis	Slope	Intercept	ANOVA Prob. (p)
Procedures (P)	828.5	18	76.3	683	824	935	-0.26	-0.83	3.34	797	0.351
Complications (C)	23.4	1.7	7.23	10	24.5	42	0.61	1.56	-0.637	29.4	0.049
Complications %	2.8%	0.22	0.92	1.3	2.75	5.3	0.88	1.86	-0.00087	0.037	0.032
Errors (E)	16.7	1.26	5.37	8	16	29	0.36	0.06	-0.55	21.9	0.019
E/C %	71.9%	2.4	10.2	53.8	75	86.4	-0.65	-0.74	-0.00545	0.771	0.252
E/P %	2.0%	0.16	0.66	1	2.05	3.6	0.47	0.65	-0.00074	0.022	0.009
Primary Index Errors	6.4	0.69	2.93	2	6	13	0.5	-0.08	-0.129	7.61	0.348
Total Index Error	10.01	1.02	4.32	4	9.5	20	0.63	0.22	-0.304	12.9	0.124
Primary Index Error/P %	0.72%	0.11	0.46	0	1	1	-1.08	-0.94	0.00001	0.0071	0.962
Total Index Error/P %	1.2%	0.12	0.53	0.4	1.15	2.5	0.68	0.48	-0.00044	0.016	0.067
Primary Index Error/C %	27.9%	2.68	11.4	11.1	27.9	50	0.49	-0.49	0.00141	0.265	0.794
Total Index Error/C %	42.9%	2.82	12	18.5	43.95	63	-0.31	-0.41	-0.00238	0.452	0.675

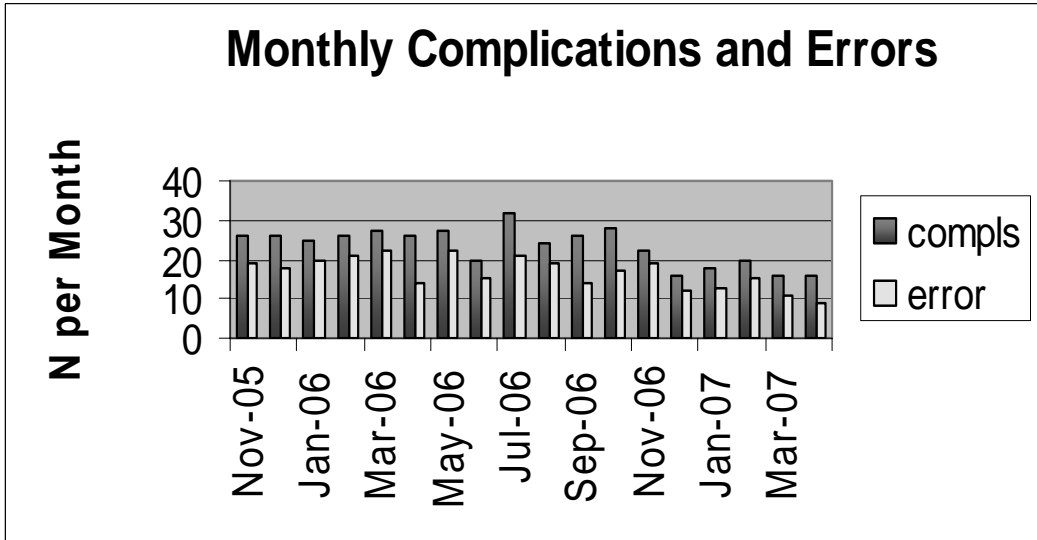
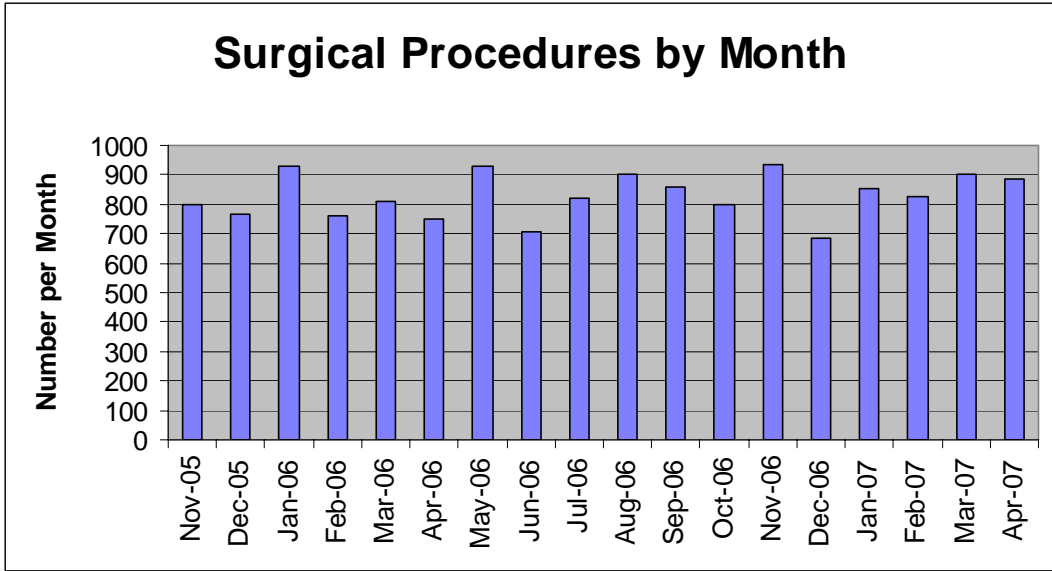


Figure 7. Surgical Procedures, Complications, and Errors by Month

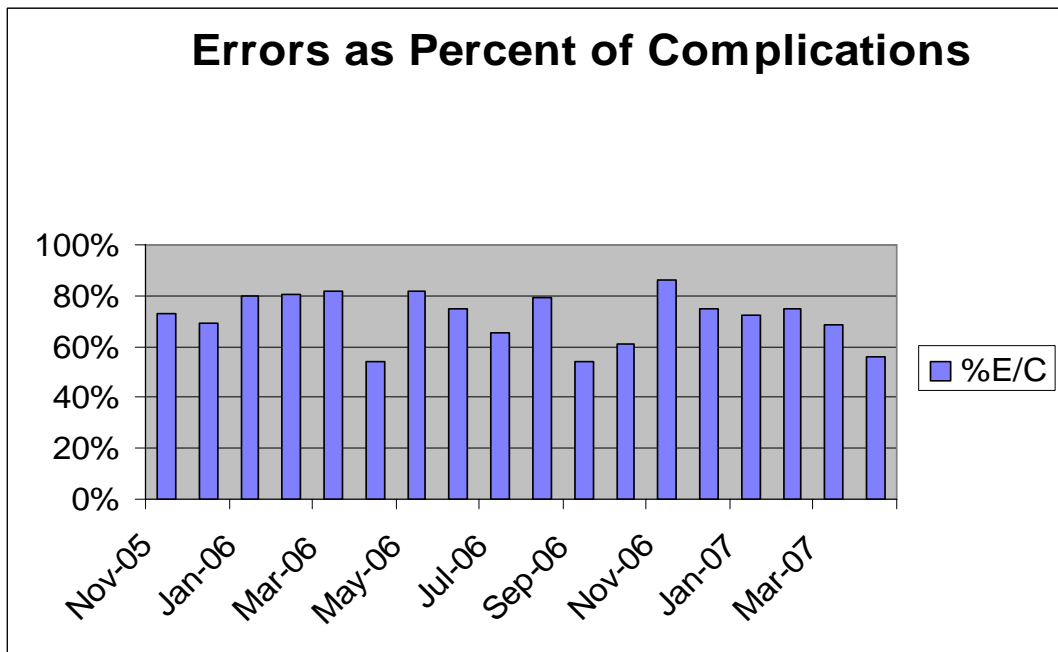
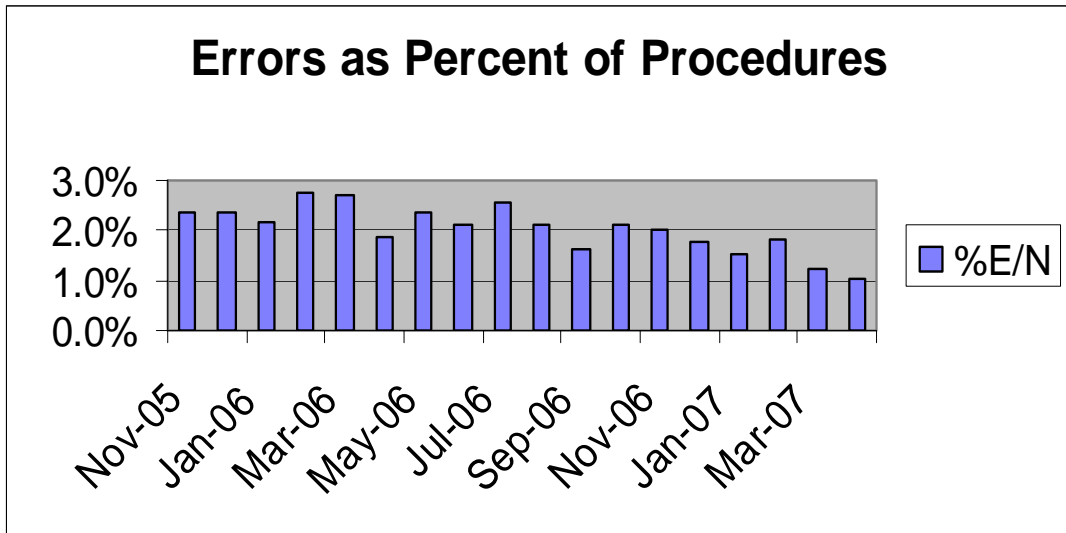


Figure 8. Errors as Percent of Procedures and Percent of Complications
 The percentage of procedures associated with a reported error (upper) and the percentage of complications reported to have been associated with an error (lower).

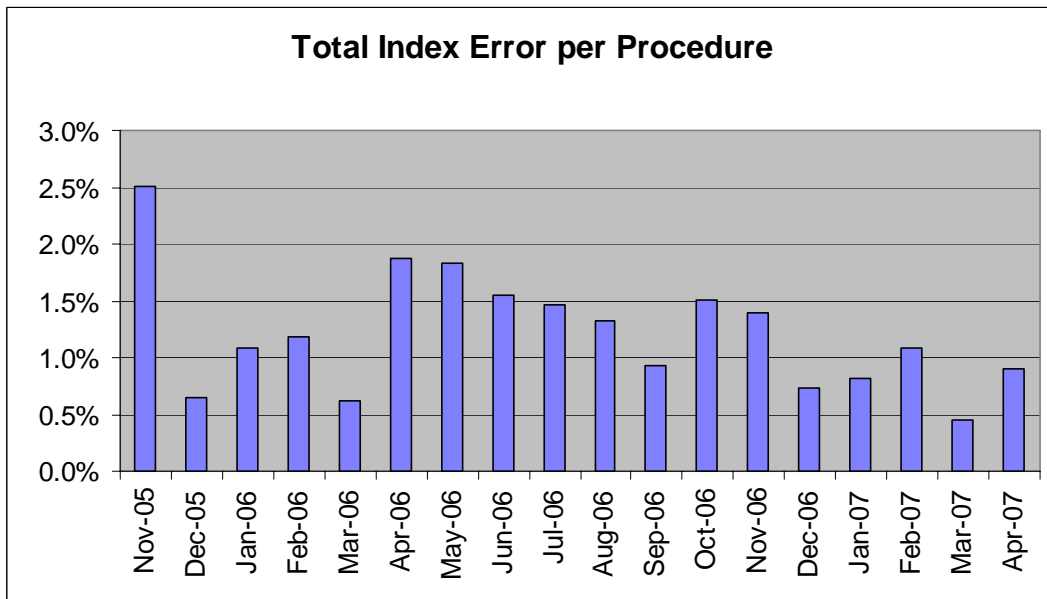
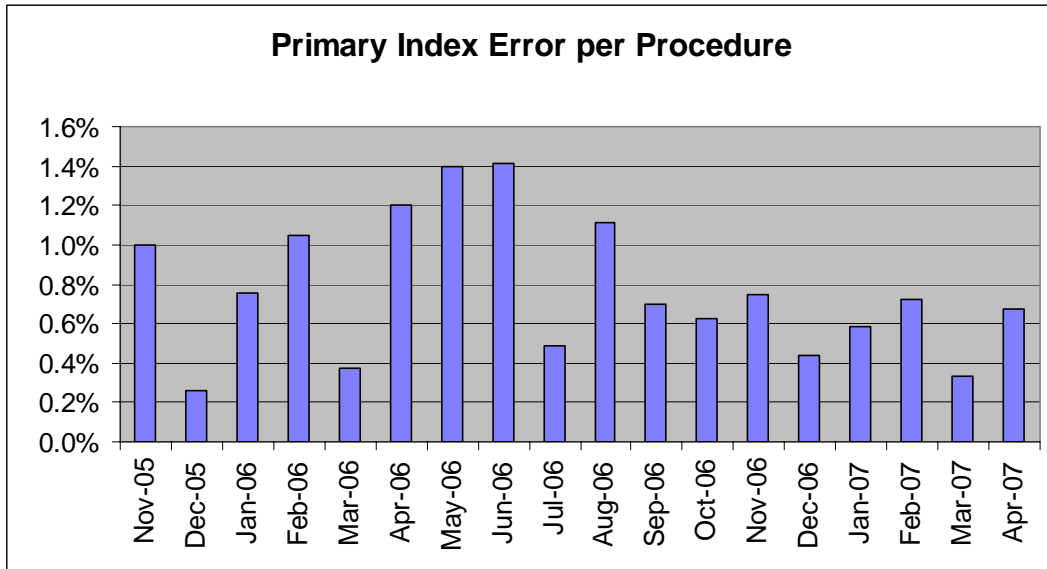


Figure 9. Index Error as Percent of Procedures

The percentage of procedures associated with an index error reported as the primary error (upper) and the percentage of procedures reported with any index error (lower).

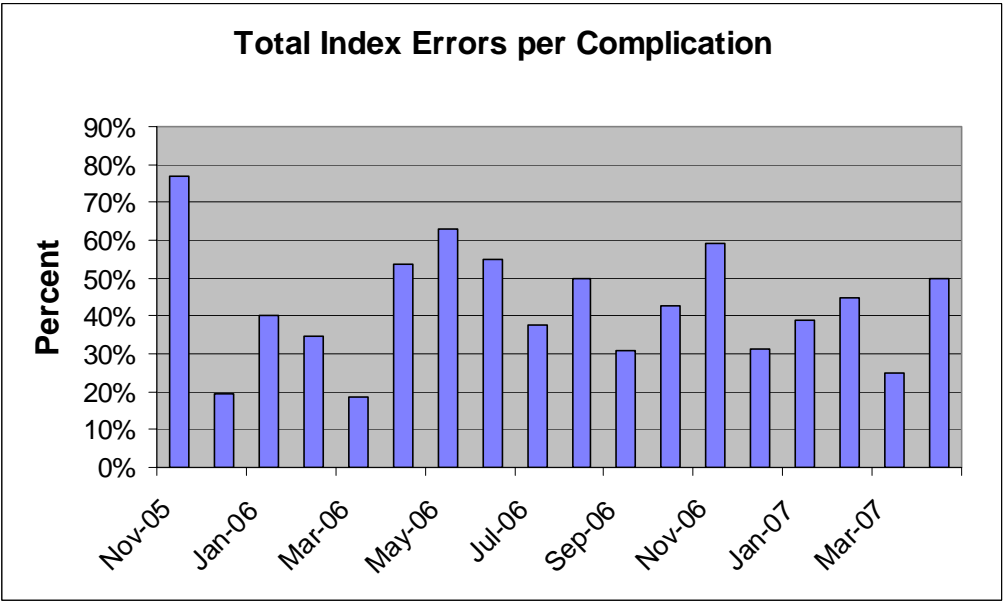
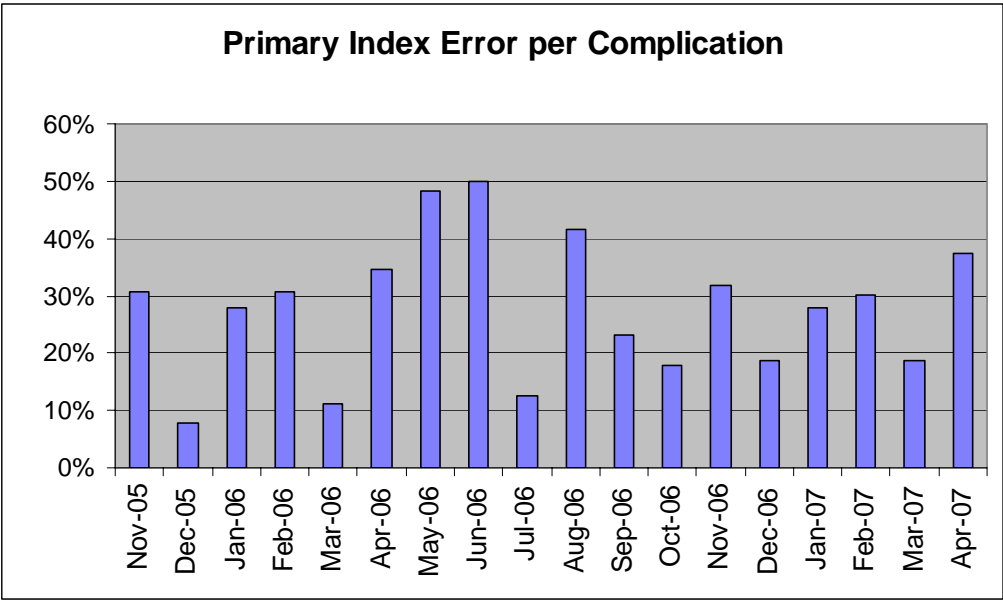


Figure 10. Index Error as Percent of Complications
 The percentage of reported complications with an index error as the primary error (upper) and the percentage of reported complications with any reported index error (lower).

There was no difference in the overall distribution of outcomes (Table 16) and no difference between the likelihood of “serious error” (Outcome Scores 3+4+5).

Table 16 . The Outcome Scores Before and After Training

Score	Outcome	Pre-Training	Post-Training
5	Death	16%	17%
4	Disability	9%	9%
3	Temp Disability	25%	22%
2	Prolonged	35%	38%
1	No Adverse Effect	15%	13%

The distribution of the impact of the error on the outcome (what percentage of the complication was due to the medical error in 25% increments) was compared before and after training by Kolmogorov Smirnov testing. There was no difference in the distribution of impact (pre-mean =63.0% impact, SD 29.2; post-mean=63.8% impact, SD 31.0; KS= .82).

Trendlines (Figure 11) were developed to assess the data over the 18 month study timeline. Significant trends were identified across the entire study period and demonstrated a progressive decrease in complication rate and error rate. The level of surgical activity was constant over the entire 18 months. Analysis of the complication rate by month for the 12 month period immediately preceding the initiation of the study showed a very consistent complication rate with no evidence of decrease over time.

Trend lines did not demonstrate a decrease in the percentage of errors attributable to the first index error reported (primary index error) or any index error reported (any error).

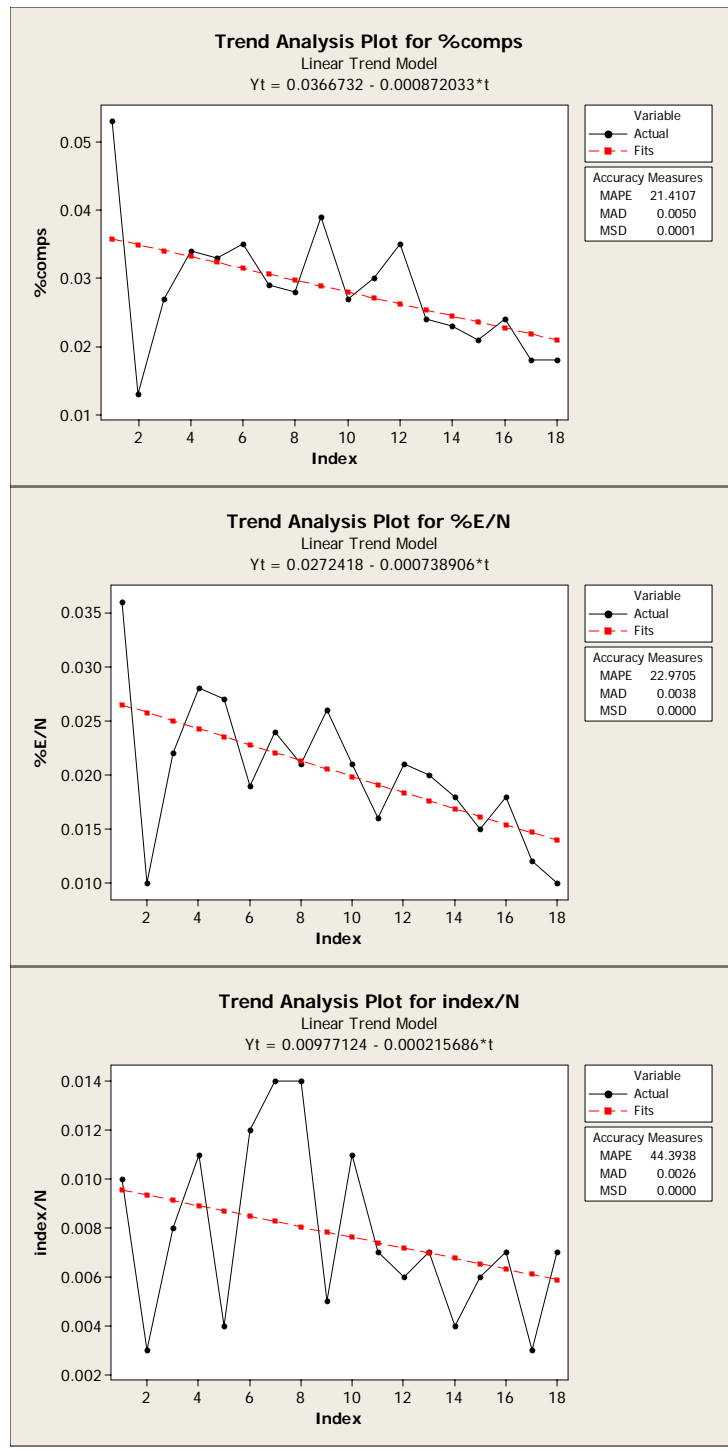


Figure 11. Trend Analysis for Complications, Errors, Index Errors
 Trend lines over the entire 18-month study period for monthly complication rate, monthly error rate, monthly index error rate, percent of complications with a primary index error, and percent of complications with any index error.

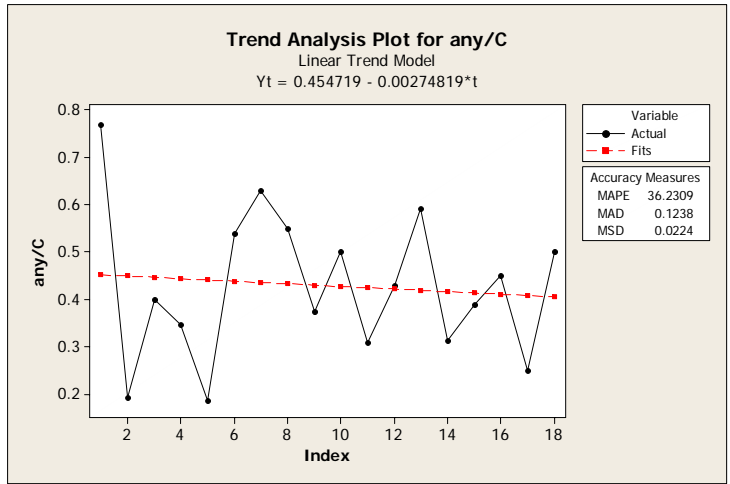
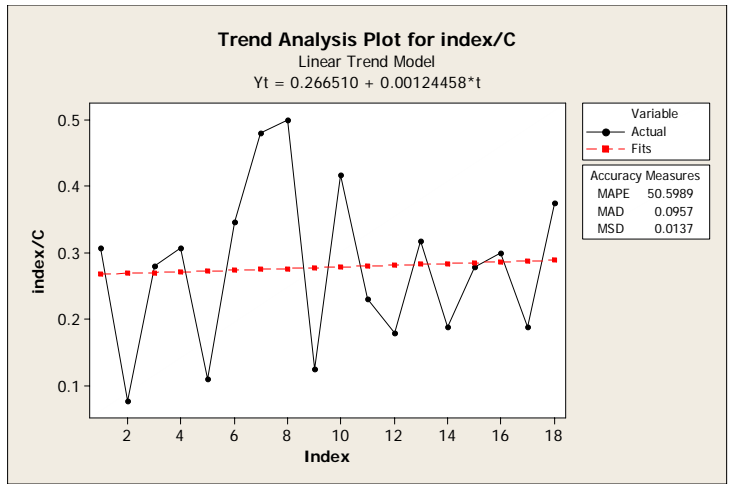


Figure 11. (Continued)

5. Discussion

Since the publication of “To Err is Human,” now almost 8 years ago, there has been a plethora of studies and articles relating to the subject of medical error. More importantly, there has been a groundswell of interest in the development of programs to train individuals and teams to minimize error¹¹²⁻¹¹⁴. However, there have been very few studies which actually look at patient outcomes in measuring whether the program is effective, and even fewer studies which define a rigorous methodology with validation and reliability testing. This is particularly true in surgery, where there have been no published studies which actually measure the impact of medical error on surgical patient outcome. This study has attempted to present a methodology to develop and validate a classification system for error, develop a training program designed to address the most common types of errors, and then measure whether the training program improved patient outcomes over the time period following the training.

5.1 Analysis of Results

It appears quite clear that the study as a whole resulted in a sizable decrease in medical errors, no matter how this was quantitatively expressed. There were fewer complications, both in absolute number and as a percentage of procedures performed. There were fewer errors, both in absolute number and as a percentage of procedures performed. Thus it appears safe to conclude that the overall program was effective in

enhancing the actual outcomes of patients undergoing major surgical procedures. It is less clear what was responsible for that impact. Analysis of the pre-training baseline data already demonstrated a significant reduction in complications over the 12 month period. Examination of the trendlines confirms that there is a fairly constant decrement in both complications and errors that would appear to antedate the actual training program. And most importantly, there did not seem to be any change in the percentage of errors that were attributable to the index errors which were the subject of the training.

The data demonstrate that there was in fact a steady state condition over the 18 months, as evidenced by a very constant number of procedures, a constant distribution of outcome scores, and a consistent contribution of error to complications. Review of the number and percent of complications during the twelve months prior to initiation of this study clearly showed that the reported complication rate was stable. Since the pilot study had taken place for several months before the actual baseline data collection commenced, one would assume that the “halo effect” of a change in reporting would most likely be past. Similarly, any “Hawthorne Effect,” a temporary change in behavior related to the fact that there was a “study,” would be expected to demonstrate an early change followed by a return to baseline. This study showed a continuous and ongoing effect over the 18 month period, no decrease after the training, and no rebound return to baseline after time had elapsed post-training. It therefore seemed surprising that there was a continued slow decrease in the numbers of complications and the numbers of errors. The absence of a discrete effect of the training on either the trend lines or on the index errors suggests that the training program itself was perhaps a minor contributor and the major factor leading to the improvement in patient outcome was attributable to the fact that everyone in the

program, resident and attending alike, was participating in error analysis on a weekly basis over a greater than 18 month period. Perhaps it was the sustained awareness of the possibility of error, due both to repetition as well as to expectation, that impacted the “culture” of the surgical program.

It is otherwise difficult to explain the seeming paradox that there was no apparent focused effect of the training on the material actually trained, yet there was a substantial decrease in the incidence of complications and errors. It is certainly true that the measurable outcome, complications, was self-reported and thus there is a possibility of inaccurate reporting. Countering this is the fact that this reporting mechanism has been in place for over two decades, is widely accepted, and undergoes weekly oversight by the chairman and the faculty of the department of surgery. Similarly, the marked consistency of the distribution of outcome scores and the attribution of impact before and after training would suggest that the likelihood of bias in data reporting is small. If there had been a change in reporting, major adverse outcomes (including disability and death) would continue to be reported, and a change in the reporting of lesser complications would alter the distribution of outcome scores or percent error. No change in outcome score was seen, suggesting that both major and minor complications continued to be reported accurately.

Another theoretical concern is that the training itself was not actually effective. Although there had been an extensive effort made and supporting data gathered to indicate that the training would be highly likely to be effective, it is certainly possible that the training was not, in fact, effective. However, this training module was subjected to

substantially more analysis and oversight than the majority of training tools used in surgical education suggesting that it was at least as effective as most comparable training exercises. It is probably not that surprising that a single, non-repeating training session, no matter how well designed and implemented, without reinforcement and without additional feedback, would have minimal transfer from the training environment to the work environment.

Is it possible that the residents were already well trained in error analysis, and thus the additional training would produce little to no effect? While this can't be summarily dismissed, there is no evidence that any training had taken place prior to this study. It would appear that residents were "naïve" to error analysis and reporting, and error analysis was not a component of either the didactic or clinical experience of the residents.

Also, it is certainly possible that the wrong group of subjects was trained. Most of the major decisions regarding surgical procedures are made by attending surgeons, not residents. Specifically, the types of errors identified as being the most common (judgment errors, incomplete understanding of the problem, carelessness/inattention to detail) are attributable mostly to the attending physician. This factor cannot be overlooked in the interpretation of the data. Until a study focused on the attending surgeon is conducted, it is not possible to say whether training the attending physicians might have had a more noticeable impact on the incidence of index errors. However, in spite of this concern, there appears to have been a significant reduction in complications and errors over the course of the study period.

This study has made a number of important observations which certainly must be verified by additional well designed trials.

Hypothesis 1. Is it possible to create a valid classification system for errors occurring in surgery? Yes, using the published literature as a basis for the classification and using experts in surgery and in patient safety, it is possible to create a useful classification system for surgical error. Although there are still traditional perceptions among surgeons on the nature and types of error as well as the “onus” of error in general, it is surprising that the agreement among a large number of academic surgeons in the absence of specific training in medical error was so substantial and consistent across three different surgical scenarios and a wide range of academic institutions.

Of perhaps greater interest is that the use of the classification system, both in the validation studies as well as in the actual data collection, showed that the overwhelming majority of errors occurring in association with surgical complications are NOT due to system problems or communication problems. This may be due to the fact that the study was not specifically designed to analyze the possibility of a “cascade” of errors. However, this study suggests that perhaps in fields of medicine where significant responsibility and direct risk of harm are in the hands of a single individual, the surgeon, the number of errors due to the surgeon him/herself may be substantially larger than the background risk of system and communication errors cited in the current literature. If this observation is true, it supports the concept that the types of errors most likely to occur may be different in different fields of medicine and therefore that there is a need for specific classification systems in different fields of medicine.

It is also important to highlight that the most “popular” surgical errors currently being discussed and reported as “sentinel events,” wrong side surgery, wrong patient surgery, wrong operation, did not occur during the entire 18 month experience. This suggests that the “extreme” errors that are reported to national and state organizations as sentinel events are a very small fraction of total error and that they may not be reflective of the types of errors occurring on a day to day basis in the conduct of surgery.

Hypothesis 2. Can this classification system demonstrate reliability when used by practicing academic physicians? Yes. As demonstrated in this study, there is a high degree of consistency among academic surgeons. There is perhaps room for additional refinement in the classification system, but overall the reliability exceeds the benchmark of 70% recommended by Wallace. One might argue that the terminology used, originally selected to be more easily understandable and interpretable by surgeons, was not sufficiently exclusive, since surgeons seemed to have difficulty deciding whether the error was due to a specific type of index error. Perhaps a different breakdown of the individual human factors might be more exclusively and consistently applied to avoid changing the data collection instrument after it was implemented. As used in this study, the data instrument was applied identically in both the pre-training and post-training setting. It was quite interesting, however, to note that surgeons had a similar level of disagreement about whether an error had occurred as they had in deciding the specific human factor which was involved in a given scenario. This variability appeared to be less within the local institution data, suggesting that there is an “institutional understanding” of what constitutes error, and that this varies from institution to institution. As additional studies are published which lead to a more “standardized”

definition of error within the field of surgery, it is likely that the variability between institutions and the variability between surgeons in their understanding of medical error will be minimized. This observation suggests that such studies must include a component of very structured training of the individuals who are going to report error.

Hypothesis 3. Can we determine a baseline incidence of the types of common errors?

Yes. The data would suggest that at the very least, the first 6 months of baseline data collection was similar to the data collected during the pilot study. In addition, even when the incidence of complications and errors was decreasing, the distribution of patient outcomes and the attribution of impact of the errors was remarkably consistent.

Hypothesis 4. Can we reliably identify the most common types of errors in an academic

surgical setting? Yes. Both the validation data and the baseline data demonstrate that surgeons in an academic setting, with minimal specific training in medical error can reliably assess the presence and type of error (as demonstrated by greater than 70% agreement, agreement between outside academic surgeons and internal ongoing data collection, and constancy within an institution over an 18 month period of data collection). The data, at least as collected, show that the majority of errors occurring in surgery are not system errors or communication errors. This may be counter to current “wisdom” in the field of error analysis, and it is likely that the outcomes would be very different in fields such as radiation oncology or anesthesiology, where the actual “action” is very dependent on equipment, where the consequences of actions are in fact very tightly “coupled” (doing “x” likely leads immediately to “y”). This observation reinforces our recommendation that classification systems for error should be specialty

specific, based on data collected in that specialty. At least as observed in this study, human factors appear to be responsible for the majority of errors in surgery, and these tend to cluster around the actions of the surgeon in the specific areas referred to as index errors: errors in judgment, incomplete understanding of the problem, and carelessness/inattention to detail.)

Hypothesis 5. Can we measure the impact of error on the actual outcomes of patients?

Yes. The validity testing among the academic surgeons demonstrated a high degree of consistency in scoring the outcome and defining the degree to which the complication was the result of error. This was further supported by the consistent responses over time in the broader clinical context of baseline and post-training data collection. In fact, the total surgical experience of a large, complex academic institution, as measured in the baseline data collection, showed remarkable stability in the decisions of the surgeons reporting the impact of error on patient outcome.

Hypothesis 6. Can we develop and validate a simulation-based training module for

resident physicians? Yes. Using the experience of a large number of academic surgeons and then reviewed by other groups of academic surgeons with special expertise in developing training materials resulted in a “training package” which was determined to be “effective.” It is of some concern that this training program did not appear to “transfer” to the work environment. The training program used in this study is at least as rigorous as the majority of training materials currently in use in graduate medical education. Further, the degree of review and analysis was substantially greater than is typically used in medical training programs. One can conclude that it is possible to

validate the training module as perceived by experts. One cannot necessarily extend that assessment to include the effectiveness of the module to transfer to the workplace and improve patient outcome.

Hypothesis 7. Can a validated training module improve surgical outcomes after a cohort of residents is trained? Uncertain. It is clear that the overall project resulted in a substantial decrease in complications and also in errors. So the *overall project* resulted in an improvement in patient outcome. But it is not clear that the actual training module was responsible for this outcome. It may be more likely that the continuous exposure to analysis of error resulted in a systemic change in attitudes and a greater awareness of error. It is not possible to conclusively answer this question at this time, although the trend line analysis clearly suggests that the change in outcome was apparent months before the training module was conducted. Most importantly, there is no evidence of any change in the types of errors specifically trained, either in a before or after, distribution free comparison or in the regression analysis with ANOVA.

Power Analysis. Using the number of surgical procedures performed as the determining factor for sample size, and not error, a power of 0.10 would require about 7000 procedures in each group. Since there were over 9000 procedures in the baseline group, and approximately 5000 procedures in the post-training group, the power of a before-after comparison is probably about 0.20. The regression or trend line of the data over the entire 18 months should be sufficiently powered to be able to identify a difference if a difference is present.

Confidence. Confidence must be considered from both the statistical sense as well as the reasonability sense. Statistically, 95% confidence intervals were utilized for all comparisons and all regressions. The Kolmogorov Smirnov test was used to assess normality of the data when parametric procedures were employed. Fisher's Exact Test was used to analyze frequency data to avoid the need to consider normality, particularly when addressing conditional probabilities and frequencies. Consequently the data analysis should be rigorous and unbiased, and therefore likely to be "accurate." Of more importance, however, is the level of confidence that the final conclusion is consistent with modern theories of training (reasonability). Particularly since our conclusions question several concepts that are currently strongly held, most notably that the errors in surgery were NOT typically system errors or communication errors and that a well designed training module did not appear to produce improvements in the quality of care, the results must also "make sense." To accomplish this, it is important to clearly distinguish between the literature in "patient safety" and the literature in "organizational training." Clearly, the training literature is emphatic that effective training requires repetition, frequent assessment with feedback, overlearning, and reinforcement, which were not a component of this training module nor is it typically included in most "courses" provided for residents. Accordingly, it is reasonable to accept with confidence that a single training module, no matter how well designed and implemented, may NOT transfer to the clinical environment at a different time and place. Similarly, it is reasonable to accept that requiring the residents to consider the possibility of error in a formal way every week over an extended period of time IS very likely to result in a sustained attention to the possibility of error, which could become incorporated into the

day to day activities of the workplace (transfer of training). Isolated training modules are certainly useful, but as an adjunct to a larger, more robust and sustained training process that will impact on the broader culture of the medical learning and practice environment.

5.2 Future Plans

This study has raised two very provocative possibilities: 1.) that in some areas of healthcare individual human factors errors may be more important causes of adverse events than system or communication error, and 2.) that well designed single episode training sessions may not transfer to the workplace to patient outcome. The most important next step is to reproduce the results. This will require conducting a similar study at a new clinical site. Although much of the validation and reliability testing will not need to be repeated, baseline data will need to be accumulated. As evidenced in this study, the collection of only 6 months of baseline data would probably have overlooked the importance of the changes which occurred during the baseline period as a consequence of the weekly data analysis and reporting. Therefore it will require a model similar to the present study. However, this study demonstrated that the percentage of major complications associated with error may be stable, suggesting that a corroborating study could use historical complication rates over a period of time, compared to similarly collected complication rates after a training experience. While such a study would not include “granular” information regarding types of errors, outcome scores, impact on patients, it could be done much more simply. Alternatively, a study could introduce a regular error analysis to the existing morbidity and mortality process, WITHOUT a

formal training module, to determine if simply having residents analyze and report errors will result in a decrease in complications and errors.

A provocative possibility for future study is training the attending physicians instead of the residents. Since many of the decisions made in the planning and execution of a surgical procedure, including preoperative and postoperative care, are either made or approved by the attending physician, training the attending may be more effective than training the resident. Whether this is culturally feasible in most institutions will determine whether such a study can be conducted.

Future studies might include several components which were not included in this study. Such additional measures might include a cascade analysis, to determine if the evident error is a component of a chain of errors; a systems analysis, to determine if the evident error is a component of a system of organized activities; and a preventability analysis, to determine if and at what stage the error could reasonably have been identified and prevented from occurring. These were not included because of the substantial advanced knowledge that would be required of the individuals submitting the weekly information, which would have biased the interpretation of any results. Inclusion of these supplementary considerations in future studies will enhance the robustness of the conclusions and facilitate the future application of patient safety training. This would presumably require a highly trained staff for data collection and a formal training program for the physicians.

An important component of error analysis is the recognition of latent errors (near miss). An effective program of near miss analysis will contribute substantially to the

identification of latent errors, before they occur. In this study, the voluntary program of near miss reporting was unsuccessful. There are many reasons to explain this. It was inconvenient. It required initiative on the part of the resident to report. It involved an interruption of workflow. It wasn't "user-friendly" in the sense of a well-designed "front end" with prompts and selections to facilitate reporting. A well-designed, easy-to-use, and readily available process that does not interrupt workflow, should allow reporting and analysis of a high volume of near misses and potentially uncover latent system errors which may contribute to or underlie the human factor errors recognized in the post-hoc analysis of surgical complications. Recognition of the most common types of near miss events could be extremely helpful in designing future training modules to minimize surgical error.

5.3 Conclusion

This study has demonstrated that it is possible to develop and validate a classification system for error in surgery. In addition, the data have demonstrated that the types of errors in surgery may be different than in other areas of medicine, reaffirming that error classification is probably domain specific, and suggesting a benchmark that studies which address medical error should focus on the types of error likely to occur and include validation of the classification system utilized. Finally, since the purpose of analyzing errors and developing training programs to enhance patient safety is to actually decrease medical errors, programs that are intended to improve patient safety should actually measure outcomes that are relevant to patient safety. This should include the

analysis of the types, frequency, and severity of error within the relevant medical content domain.

This study has demonstrated that the process of measuring error on a regular basis as a component of morbidity and mortality analysis leads to a sustained reduction in complications presumably by decreasing the errors which appear to be tightly linked to them. It also suggests that a one-time, focused training program which addresses specific, common types of errors may have minimal effect on patient outcomes. Recurring training with feedback and reinforcement may rectify this apparent limitation.

Most importantly, this study supports the concept that it is the culture of safety which will influence patient outcome, more than a specific training program. To the degree that training modules consistently maintain and augment this culture of safety, they will be an important component. Emphasis on the training module rather than on the culture may have limited effect on the goal of this endeavor, safer surgical care for patients.

References

1. Committee on Quality of Health Care in America, the Institute of Medicine. *To Err Is Human, Building a Better Health System*. National Academy Press; 2000.
2. Bagian JP. Patient safety: lessons learned. *Pediatr Radiol* 2006; 36(4):287-90.
3. *Patient Safety and Quality Improvement Act of 2005*. Public Law 109-41, 2005.
4. McNutt RA, Abrams R, Arons DC. Patient safety efforts should focus on medical errors. *JAMA* 2002; 287(15):1997-2001.
5. Clinton HR, Obama B. Making patient safety the centerpiece of medical liability reform. *N Engl J Med* 2006; 354(21):2205-8.
6. Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. *JAMA* 2002; 288(4):501-7.
7. Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? *JAMA* 2005; 293(19):2384-90.
8. Lee TH. A broader concept of medical errors. *N Engl J Med* 2002; 347(24):1965-7.
9. Cooper JB, Gaba DM, Liang B, Woods D, Blum LN. The National Patient Safety Foundation agenda for research and development in patient safety. *J. Gen Med* 2000; 2(3):E38.
10. Fassett WE. The Patient Safety and Quality Improvement Act of 2005. *The Annals of Pharmacotherapy* 2006; 40(5):917-24.
11. Leape L. Error in Medicine. *JAMA* 1994; 272(23):1851-7.
12. Leape LL, Bates DW, Cullen DJ, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA* 1995; 274(1):35-43.
13. Leape LL. A systems analysis approach to medical error. *J Eval Clin Pract* 1997; 3(3):213-22.

14. Gaba D. *Human Error in Dynamic Medical Domains*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1994.
15. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery* 1999; 126(1):66-75.
16. Senders JW, Neville PM. *Human Error: Cause, Prediction, and Reduction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1991.
17. Bogner MS. *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1994.
18. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. *Med Care* 1989; 27(12):1148-58.
19. Cooper JB, Newbower RS, Kitz RJ. An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection. *Anesthesiology* 1984; 60(1):34-42.
20. Mold JW, Stein HF. The cascade effect in the clinical care of patients. *N Engl J Med* 1986; 314(8):512-4.
21. Allnut MF. Human Factors in accidents. *Br J Anaesth* 1987; 59:856-64.
22. Bosk CL. *Forgive and remember: Managing Medical Failure* 2nd Ed. ed. Chicago: University of Chicago Press; 2003.
23. Reason J. Understanding adverse events: human factors. *Qual Health Care* 1995; 4(2):80-9.
24. Vincent C. Understanding and responding to adverse events. *N Engl J Med* 2003; 348(11):1051-6.
25. Anderson JG, Jay SJ, Anderson M, Hunt TJ. Evaluating the Capability of Information Technology To Prevent Adverse Drug Events: A Computer Simulation Approach. *J Am Med Inform Assoc* 2002; 10(3):290-1.
26. Anonymous. Strategies for preventing med errors. *OR Manager* 2006; 22(3):11, 3.
27. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. *J Am Med Inform Assoc* 2001; 8(4):299-308.
28. Bates DW, Leape LL. Doing better with critical test results. *Jt Comm J Qual Patient Saf* 2005; 31(2):66-71.

29. Tamuz M, Thomas EJ. Classifying and interpreting threats to patient safety in hospitals: insights from aviation. *J Organiz Behav* 2006; 27:919-40.
30. Vincent C. *Patient Safety*. London: Elsevier; 2006.
31. Van Cott H, ed. *Human Errors: Their Causes and Reduction*; 1994.
32. Hamman WR. The complexity of team training: what we have learned from aviation and its applications to medicine. *Qual Saf Health Care* 2004; 13 Suppl 1:i72-9.
33. Cook RI, Woods DD. Adapting to new technology in the operating room. *Hum Factors* 1996;38(4):593-613.
34. Cuschieri A. Reducing errors in the operating room: surgical proficiency and quality assurance of execution. *Surg Endosc* 2005; 19(8):1022-7.
35. Dain S. Management of high-risk perioperative systems. *Semin Cardiothorac Vasc Anesth* 2006;10(2):180-3.
36. Dankelman J, Grimbergen CA. Systems approach to reduce errors in surgery. *Surg Endosc* 2005; 19(8):1017-21.
37. Halvorsrud R, Hagen S, Fagernes S, Mjelstad S, Romundstad L. Trauma team training in a distributed virtual emergency room. *Stud Health Technol Inform* 2003; 94:100-2.
38. Hales BM, Pronovost PJ. The checklist--a tool for error management and performance improvement. *J Crit Care* 2006; 21(3):231-5.
39. Nolan TW. System changes to improve patient safety. *BMJ* 2000; 320(7237): 771-3.
40. Reason J. *Human Error*. Cambridge: Cambridge University Press; 1990.
41. Helmreich RL SH, ed. *Team Performance in the Operating Room*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1994.
42. Cook R, O'Connor M, Render M, Woods, D. *Operating at the Sharp End: The Human Factors of Complex Technical Work and Its Implications for Patient Safety*: American College of Surgeons; 2004.
43. Plato KR, Lee D. *The Republic*. London, England: Penguin Classics; 2003.
44. Jastrow JN, 1936. . *The Story of Human Error*. New York: D. Appleton-Century Company; 1936.

45. Reason JT, Mycielska MK. *Absent-Minded? The Psychology of Mental Lapses and Everyday Errors*. Englewood Cliffs, NJ: Prentice-Hall; 1982.
46. Neuhauser D, Ernest Amory Codman MD. *Quality and Safety in Health Care* 2002; 11:104-5.
47. Jastrow J. *Managing Your Mind*. New York: Grenlenger; 1931.
48. Newell A, Simon H. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall; 1972.
49. Rouse W. *Models of human problem solving: Detection, diagnosis and compensation for system failures*. Baden-Baden, FRG; 1981 September 1981.
50. Rasmussen J. *Human errors: a taxonomy for describing human malfunction in industrial installations*. *Journal of Occupational Accidents* 1982; 4:311-35.
51. Rasmussen J. *The role of error in organizing behaviour*. 1990. *Qual Saf Health Care* 2003; 12(5):377-83; discussion 83-5.
52. Wallace B, Ross A. *Beyond Human Error: Taxonomies and Safety Science*. New York: Taylor & Francis; 2006.
53. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. *Qual Saf Health Care* 1991; 13(2):145-51; discussion 51-2.
54. Leape L, Brennan A, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. *New England Journal of Medicine* 1991; 324:377-84.
55. Bogner MS. Implications for patient safety. *Biomed Instrum Technol* 2006; 40(1):61-3.
56. Reason J. Too little too late: A commentary on accident and incident reporting systems. In: Lucas DA, ed. *Near Miss Reporting as a Safety Tool*: Butterworth-Heinemann; 1991.
57. Reason J. Human error: models and management. *West J Med* 2000; (6):393-6.
58. Zhao B, Olivera F. Error Reporting in Organizations. *Academy of Management Review* 2006; 31(4):1012-30.
59. Makeham MA, Dovey SM, County M, Kidd MR. An international taxonomy for errors in general practice: a pilot study. *Med J Aust* 2002; 177(2):68-72.

60. Rogers S. A structured approach for the investigation of clinical incidents in health care: application in a general practice setting. *Br J Gen Pract* 2002; 52 Suppl:S30-2.
61. Dovey SM, Meyers DS, Phillips RL, Jr., et al. A preliminary taxonomy of medical errors in family practice. *Qual Saf Health Care* 2002; 11(3):233-8.
62. Rubin G, George A, Chinn DJ, Richardson C. Errors in General Practice: Development Of An Error Classification and Pilot Study Of A Method For Detecting Errors. *Qual Saf Health Care* 2003; 12(6):443-7.
63. Leach DC. Evaluation of Competency: An ACGME Perspective. *American Journal of Physical Medicine and Rehabilitation* 2000; 79(5):487-5.
64. Leach DC. The ACGME Competencies: Substance or Form? Accreditation Council on Graduate Medical Education. *Journal of the American College of Surgeons* 2001; 192(3):396-8.
65. Wayne DB BJ, Siddall VJ, et al. Simulation Based Training of Internal Medicine Residents in Advanced Cardiac Life Support Protocols: a Randomized Trial. *Teaching and Learning in Medicine* 2005; 17:202-8.
66. Scott DJ. Patient Safety, Competency, and the Future of Surgical Simulation. *Simulation in Healthcare* 2006; 1(3):164-70.
67. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005; 190(1):107-13.
68. Rosen J, Solazzo M, Hannaford B, Sinanan M. Objective laparoscopic skills assessments of surgical residents using Hidden Markov Models based on haptic information and tool/tissue interactions. *Stud Health Technol Inform* 2001; 81:417-23.
69. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. *J Am Coll Surg* 2006; 202(5):779-87.
70. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Ann Surg* 2004; 240(3):518-25; discussion 25-8.
71. Aggarwal R, Darzi A. Technical-skills training in the 21st century. *N Engl J Med* 2006; 355(25):2695-6.
72. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84(2):273-8.
73. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 2004; 135(1):21-7.

74. Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment for laparoscopic surgery. *Arch Surg* 1997; 132(2):200-4.
75. Blackburn T, Sadler C. The role of human patient simulators in health-care training. *Hosp Med* 2003; 64(11):677-81.
76. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med* 2006; 355(25):2664-9.
77. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997; 173(3):226-30.
78. Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care* 2004; 13 Suppl 1:i2-10.
79. Schwid HA RG, Michalowski P, Ross BK. Screen-Based Anesthesia Simulation With Debriefing Improves Performance in a Mannequin-Based Anesthesia Simulator. *Teaching and Learning in Medicine* 2001; 13:92-6.
80. Bann S, Datta V, Khan M, Darzi A. The surgical error examination is a novel method for objective technical knowledge assessment. *Am J Surg* 2003; 185(6):507-11.
81. Satish U, Streufert S. Value of a cognitive simulation in medicine: towards optimizing decision making performance of healthcare personnel. *Qual Saf Health Care* 2002; 11(2):163-7.
82. Leipzig RM, Hyer K, Ek K, et al. Attitudes toward working on interdisciplinary healthcare teams: a comparison by discipline. *J Am Geriatr Soc* 2002; 50(6):1141-8.
83. Brewster LP, Risucci DA, Joehl RJ, et al. Management of adverse surgical events: a structured education module for residents. *Am J Surg* 2005; 190(5):687-90.
84. Dreyfus H, Dreyfus S. *Mind over Machine*. New York: Free Press; 1986.
85. Coyle YM, Mercer SQ, Murphy-Cullen CL, Schneider GW, Hynan LS. Effectiveness of a graduate medical education program for improving medical event reporting attitude and behavior. *Qual Saf Health Care* 2005; 14(5):383-8.
86. Volpp KG, Grande D. Residents' suggestions for reducing errors in teaching hospitals. *N Engl J Med* 2003; 348(9):851-5.
87. Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? *Qual Saf Health Care* 2003; 12(3):221-6; discussion 7-8.

88. Jagsi R, Kitch BT, Weinstein DF, Campbell EG, Hutter M, Weissman JS. Residents report on adverse events and their causes. *Arch Intern Med* 2005; 165(22):2607-13.
89. Hoff TJ, Pohl H, Bartfield J. Teaching but not learning: how medical residency programs handle errors. *J Organiz Behav* 2006; 27:869-96.
90. Kopec D TS, Levy K, Eckhardt R, Shagas G. *The State of the Art in the Reduction of Medical Errors: IOS Press; 2006.*
91. Tamuz M, Thomas EJ, Franchois KE. Defining and classifying medical error: lessons for patient safety reporting systems. *Qual Saf Health Care* 2004; 13(1):13-20.
92. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. *J Biomed Inform* 2004; 37(3):193-204.
93. Vincent C, Taylor-Adams S, Chapman EJ, et al. How to investigate and analyse clinical incidents: clinical risk unit and association of litigation and risk management protocol. *BMJ* 2000; 320(7237):777-81.
94. Vincent C, Taylor-Adams S, Stanhope N. Framework for analysing risk and safety in clinical medicine. *BMJ* 1998; 316(7138):1154-7.
95. Busse DK, Wright DJ. Classification and analysis of incidents in complex medical environments. *Top Health Inf Manage* 2000; 20(4):1-11.
96. Runciman WB, Helps SC, Sexton EJ, Malpass A. A classification for incidents and accidents in the health-care system. *J Qual Clin Pract* 1998; 18(3):199-211.
97. Foucar E. Classification of error in anatomic pathology: a proposal for an evidence-based standard. *Semin Diagn Pathol* 2005; 22(2):139-46.
98. Elder NC, Dovey SM. Classification of medical errors and preventable adverse events in primary care: a synthesis of the literature. *J Fam Pract* 2002; 51(11):927-32.
99. Benner P, Sheets V, Uris P, Malloch K, Schwed K, Jamison D. Individual, practice, and system causes of errors in nursing: a taxonomy. *J Nurs Adm* 2002; 32(10):509-23.
100. Zhang J, Patel VL, Johnson T, Shortliffe EH. Toward An Action Based Taxonomy of Human Errors in Medicine. In: Conference of Cognitive Science Society.
101. Avery AJ. Classifying and identifying errors. *Qual Saf Health Care* 2003; 12(6):404.

102. Hofer TP, Hayward RA. Are bad outcomes from questionable clinical decisions preventable medical errors? A case of cascade iatrogenesis. *Ann Intern Med* 2002; 137(5 Part 1):327-33.
103. American Academy of Family Physicians rGCfPSiFPaPC. Toxic cascades: a comprehensive way to think about medical errors. *Am Fam Physician* 2001; 63(5):847.
104. Woolf SH, Kuzel AJ, Dovey SM, Phillips RL, Jr. A string of mistakes: the importance of cascade analysis in describing, counting, and preventing medical errors. *Ann Fam Med* 2004; 2(4):317-26.
105. Chang A, Schyve PM, Croteau RJ, O'Leary DS, Loeb JM. The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. *Int J Qual Health Care* 2005; 17(2):95-105.
106. Noe R. *Employing Training and Development*, 3rd Edition. McGraw Hill Irwin; 2005.
107. Barnett SM, Ceci SJ. When and where do we apply what we learn a taxonomy for far transfer. *Psychological Bulletin* 2002; 128(4):612-37.
108. Ford JK, Quinones MA, Segoo DJ, Sorra JS. Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology* 1992; 45.
109. Tracey JB, Tannenbaum SI, Kavanagh MJ. Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology* 1995; 80(2):239-52.
110. Wexley KN, Baldwin TT. Posttraining strategies for facilitation positive transfer: An empirical exploration. *Academy of Management Journal* 1988; 28(3):503-20.
111. Brannick M, Levine E. *Job Analysis: Methods, Research, and Applications for Human Resource Management in the New Millenium*. Thousand Oaks, California: Sage Publications; 2002.
112. Blum RH, Raemer DB, Carroll JS, Sunder N, Felstein DM, Cooper JB. Crisis resource management training for an anaesthesia faculty: a new approach to continuing education. *Med Educ* 2004; 38(1):45-55.
113. Blum RH, Raemer DB, Carroll JS, Dufresne RL, Cooper JB. A method for measuring the effectiveness of simulation-based team training for improving communication skills. *Anesth Analg* 2005; 100(5):1375-80.
114. Cole KD, *Gerontol Geriatr Educ* 2003;24(2):35-49.

Bibliography

1. Armitage P, Berry G, Matthews JNS, *Statistical Methods in Medical Research* 4th Edition, Blackwell Science, 2002.
2. Banja.JD *Medical Errors and Medical Narcissism*. Jones and Bartlett Publishers, Sudbury, Mass. 2005.
3. Bogner, MS editor. *Human Error in Medicine*, Lawrence Erlbaum Associates, Hillsdale, NJ. 1994.
4. Brannick MT. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Lawrence Erlbaum Assoc. Mahway, NJ, 1997.
5. Brannick MT, Levine, EL. *Job Analysis-Methods, Research, and Applications for Human Resource Management in the New Millenium*. Sage Publications Thousand Oaks. 2002.. “Test Validation”.
6. Committee on Quality of Health Care in America, Institute of Medicine. *Crossing the Quality Chasm,- a new health system for the 21st century*. National Academy Press, 2001.
7. Cook, RI, O’Connor M, Render M, Woods, D. *Operating at the Sharp End: The Human Factors of Complex Technical Work and Its Implications for Patient Safety*. In Manuel BM, Nora PF (eds). *Surgical Patient Safety: Essential Information for Surgeons in Today’s Environment*. American College of Surgeons, Chicago. 2004.
8. Denzin NK, Lincoln, YS Editors. *Handbook of Qualitative Research*, Sage 1994.
9. Institute of Medicine of the National Academies. *Patient Safety-Achieving a New Standard for Care*. The National Academy Press 2004.
10. Jastrow, J. *The Story of Human Error*. D. Appleton-Century Company, NY, 1936. p.16-18, 389-411.
11. Kohn LT, Corrigan JM, Donaldson MS, editors. Committee on Quality of Health Care in America, Institute of Medicine. *To Err Is Human, Building a Better Health System*. National Academy Press, 2000.

12. Manuel BM, Nora PF. *Surgical Patient Safety: Essential Information for Surgeons in Today's Environment*. American College of Surgeons, 2004.
13. Montgomery DC, *Design and Analysis of Experiments*, 6th Edition. John Wiley and Sons, 2005.
14. Munro BH, *Statistical Methods for Health Care Research*, 5th Edition, Lippincott Williams and Wilkins 2005.
15. Noe RA, *Employee Training and Development*, 3rd Edition. McGraw Hill Irwin, 2005.
16. Peters, GA., Peters BJ. *Human error: causes and control* Taylor and Francis, Boca Raton 2006.
17. Reason, J. *Human Error*. Cambridge University Press. Cambridge. 1990.
18. Salas E, Bowers CA, Edens E. *Improving Teamwork in Organizations: Applications of Resource Management Training*. Lawrence Erlbaum Assoc. Mahwah, NJ, 2001.
19. Senders JW, Moray, NP. *Human Error- Cause, Prediction, and Reduction* . Lawrence Erlbaum Assoc. Hillsdale , NJ 1991.
20. Snedecor GW, Cochran WG. *Statistical Methods*. Iowa State University Press, 1989.
21. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. Third Ed. Pearson Education, 2007.
22. Vincent, Charles. *Patient Safety*. Churchill Livingstone-Elsevier. 2006.
23. Yin RK *Case Study Research, Design and Methods*. Second Edition. Applied Social Research Methods Series Volume 5. Sage Publications 1994.
24. Wallace B, Ross, A. *Beyond Human Error-Taxonomies and Safety Science*. CRC Taylor and Francis Group, NY, 2006.
25. Wickens CD, Lee JD, Liu Y, Gordon Becker, SE. *An Introduction to Human Factors Engineering*, 2nd edition. Pearson, Prentice-Hall 2004.
26. Woods, D, Cook R. *Perspectives on Human Error: Hindsight Biases and Local Rationality*. In Durso RS et al, eds *Handbook of Applied Cognition*. Wiley, New York. 141-171, 1999.

27. Wreathall, J. Nemeth, C. *Assessing Risk: The Role of Probabilistic Risk Assessment (PRA) in Patient Safety Improvement. Quality and Safety in Healthcare* 13:206-212, 2004.

Appendices

Appendix 1. Letter to Expert Evaluators for External Validation of Scoring Template

Dear Doctor,

I recently made the decision to pursue a PhD in Industrial Engineering. I know it sounds crazy, but I believe it will help me to address the systems and processes problems which plague modern healthcare and, more particularly, graduate medical education. My dissertation project involves the intersection of three important themes: GME, patient safety, and medical simulation. In order to proceed, I need to validate a scoring form which I designed for morbidity and mortality conference. The intended use is that each complication gets evaluated to determine if there was a medical error, some characteristics of the medical error, the actual type of error involved, and the percentage of the complication that seems attributable to the error (as opposed to the disease).

As an expert in the field of academic surgery, I would like you to take just a few minutes and complete a scoring template on three very brief scenarios. In each case, there clearly was a complication. I would like you to determine:

1. what was the complication (just a few words)
2. if the complication was due to a medical error
3. if the medical error occurred during evaluation or execution
4. if the error was a “slip” (doing the right thing but incorrectly)
5. if the error was a “mistake” (doing the wrong thing)
6. what types of error (can be more than one) are involved, by placing a “1” in the appropriate box for the most important, a “2” for the next most important, etc.
7. which of the ACGME core competencies appears to be involved
8. what percentage of the complication can be attributed to the error (in multiples of 25%)

Your response is anonymous. I will only track two pieces of information: what institution you are from (by a code number, not name) and how many years since you finished your general surgery residency (provided by you).

I am including the three scenarios, a glossary of terms, and three scoring templates. You can return them in the included envelope or by FAX to

Thank you in advance for your help. I believe this project will lead to a significant improvement in how we teach residents and, ultimately, how we provide surgical care to our patients.

Sincerely,

Peter J. Fabri M.D.

Appendix 2. Mock-up of Scoring Template with Instructions

INSTRUCTIONS

1	Patient ID	Reviewer	SCORE=
	Brief description of complication:		
	Is the complication due to a medical error? Yes <input type="checkbox"/> No <input type="checkbox"/> **If no, you are done!		
	Did the error occur during "evaluation" <input type="checkbox"/> or during "execution" <input type="checkbox"/>		
	Is the error due to a "slip" <input type="checkbox"/> (slip = doing the correct thing, incorrectly)		
	Is the error due to a "mistake" <input type="checkbox"/> (mistake is doing the wrong thing, but correctly)		
2	Type Of Error	ACGME CORE Competency	
	equipment failure (mechanical error)	Delivery of Patient Care	
	health system error (organizational error)	Insufficient or Inaccurate Medical Knowledge	
	Incomplete understanding of problem	Adult Learning; Problem Solving	
	failure to use established protocol	Interpersonal and Communication Skills	
	carelessness/inattention to detail	Professional Behavior	
	error in diagnosis	Knowing the Health Care System	
	communication error		
	judgment error		
	delay error		
	error of omission		
	technique error		
3	% of adverse outcome attributable to "error" (25,50,75,100%) :		
	Instructions:		
	1 place a number in the scoring box at the top grading the severity of the error from 1-5.		
	5= death		
	4= partially recoverable injury resulting in probable permanent disability		
	3= fully recoverable injury but with substantial temporary disability;		
	2= definite injury but without any apparent disability		
	1= definite complication but which did not adversely affect patient		
	2 Place a "1" in the template box that represents the primary medical error involved.		
	Place a "2", "3", etc in additional boxes if more than one error type. Repeat for ACGME competency.		
	3 Mark % of outcome attributable to error		

SCORING INSTRUCTIONS

Figure 12. Error Template as Sent to Evaluators

Appendix 3. Glossary of Terms Used in the Error Template and Instructions

Glossary of Terms Used in Error Analysis

ERROR- a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, not by chance.

VIOLATION- a generic term to describe actions that are wrong, that violate “the law”. For example, running a red light.

ADVERSE EVENT- an error in which patient harm has occurred.

NEAR MISS- an error which is “caught” before any harm has occurred. This could be that the event was planned or even initiated but not actually carried out, or it could be carried out and the adverse effect was prevented from occurring.

SLIPS AND LAPSES are errors which result from some failure in the execution and/or storage of an action sequence. (doing perhaps the correct thing, but doing it incorrectly).

MISTAKES are deficiencies or failures in the judgmental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it (doing the wrong thing).

ERRORS typically occur either during evaluation/planning or during execution. Errors occurring during evaluation/planning are generally “mistakes”. Errors occurring during execution are generally “slips”. Errors occurring during information processing are generally “lapses”.

ERROR TYPES. Errors are usually divided into three main categories, organizational, technical, and human.

HEALTH SYSTEM ERROR (Organizational or System Error). This category can include any error due to failure of the “system” in getting things done. It could be an error of commission (this is how the organization does things) or omission (we don’t have a procedure to prevent that.)

TECHNICAL/MECHANICAL ERROR. In healthcare, the term technical is often associated with “technique”. To avoid confusion, we consider a “technical error” to be a human error and not a mechanical malfunction. Thus there are organizational (health system), mechanical, and human errors.

HUMAN FACTORS. There are many classification systems for error caused by how humans function as individuals. The classification system is often specific to the type of activity being performed and uses language comfortable to the individuals working in the area. In surgery, we have identified the following workable classification system:

Appendix 3. (Continued)

Incomplete understanding of the problem – making a decision or committing an action based on incomplete or inaccurate information. This is an error of cognition.

Failure to follow an existing protocol- not using a protocol where one exists and is known.

Inattention to detail- usually misinterpreting subtle findings. This is usually an error due to haste or carelessness.

Error in diagnosis- This specifically relates to when a diagnosis has been made and an action taken, but the diagnosis was formally incorrect.

Failure in communication- This type of error occurs when there is a communication breakdown between human beings that leads to someone committing an error.

Error in judgment- this occurs when an individual violates a “rule”, but has correct information.

Delay in diagnosis/treatment. This error occurs when an excessive amount of time has elapsed, leading to progression of disease or a complication of disease.

Error of omission- This error occurs when someone fails to do something indicated.

Technical error- This error occurs in the process of a procedure requiring technical skill, in which the procedure is not done sufficiently well. This could be due to insufficient understanding of what to do or just not doing it well enough.

INCOMPLETE UNDERSTANDING OF PROBLEM. This is NOT the same as error in diagnosis, although it could be closely related. For example, if a patient on a ventilator is hypoventilating and having respiratory difficulty, you don't figure out that the tube is down the main stem bronchus, and the patient arrests, that's incomplete understanding of problem. But if you put in a chest tube, that's an error in diagnosis. In the first case, you made the correct diagnosis but didn't figure out the exact mechanism. If it took you four hours to figure this out, it would ALSO be a delay error. So this error term means you didn't get all the way to the end of the diagnostic tree.

FAILURE TO USE ESTABLISHED PROTOCOL An example of this could be in a code, in a trauma resuscitation, etc, if you fail to use the ACLS or ATLS protocol. This should be used when there is actually a protocol that everybody is aware of. NOT just this isn't the way we usually do it! It could also be a chemotherapy protocol, or a dialysis protocol, or a TPN protocol.

CARELESSNESS/INATTENTION TO DETAIL This error code should be used when there really was carelessness or inattention to detail and not a scapegoat term for another type of error (e.g. communication error, judgment error) although the two could certainly coexist. This occurs when somebody really isn't paying attention. For example, if you pack the small bowel under a towel and hold it out of the field with a wide retractor, only to find out an hour later that a segment of the small bowel infarcted because of blocked

Appendix 3. (Continued)

blood supply. Or you are coming across the transverse mesocolon and divide the superior mesenteric artery instead of the middle colic. (this could also be a technique error). Or taking out a segment of the colon that turns out not to include the pathology.

ERROR IN DIAGNOSIS This term applies when you treat the patient for disease X and the patient has disease Y. In the simplest case, taking out a normal appendix is an error in diagnosis. As a more complex example, opening a chest for cardiac tamponade in the ICU in a patient with a tension pneumothorax is an error in diagnosis.

COMMUNICATION ERROR This could include a botched verbal order, as well as a failure to communicate. Make sure the actual error is related to the communication and not to something else.

JUDGMENT ERROR This error type should be assigned when it is clear that someone with more experience wouldn't have made this mistake. This could also be a secondary type of error combined with another type.

DELAY ERROR This type of error should be considered when there is clearly a temporal delay which contributed to the patient outcome. If the patient outcome was not affected by the delay, it probably isn't a delay error.

ERROR OF OMISSION This occurs when something important isn't done. If it is part of a written protocol, it would be a protocol failure. Not getting an appropriate imaging study in a patient with an aortic aneurysm and hypertension and the patient developing postoperative renal failure might be an error of omission. Or failure to use DVT prophylaxis in a situation where it would be considered to be important (malignancy, obesity, long procedure, prior DVT, etc.)

TECHNIQUE ERROR I think we all know this when we see it. Cutting the common duct (although it could also be carelessness), obstructing the renal artery on an aortic aneurysm repair, etc. Use this when the complication is actually do to something that was done to the patient which, if done correctly, would have avoided the complication.

Appendix 4. Instructions for Using the Error Evaluation Template

The instructions for completing the template are straightforward. The sheet is divided into three areas, numbered 1, 2, 3. At the bottom is an abbreviated set of instructions for reference.

Section 1 is a set of simple statements. For “Patient ID”, indicate the scenario number, 1, 2, or 3. Leave “Reviewer” blank. Enter a number from 1-5 from the instructions at the bottom to indicate your assessment of the severity of the complication.

Please note that a “1” means that there was definitely a complication (e.g. enterotomy, arterial injury, the wrong medication but without adverse effect, etc.) but there was *no adverse effect* on the patient noted.

A “2” means that the patient was affected adversely, probably lengthening hospital stay, but this didn’t translate into a significant impact on the patient and no disability (minor wound infection, cystitis treated with oral antibiotics, etc.)

A “3” means when the patient left the hospital, there was still some definite disability (major wound affecting mobility, swollen leg, still needed oxygen, etc. A “4” means the patient will almost certainly have a permanent disability.

A “5” means the patient died, period, regardless of cause.

Next, answer the simple questions. First, yes, there was a complication. Otherwise we wouldn’t be here. Was there an error? If no, you don’t need to do any more. If yes, was the error in evaluation/assessment/planning or was it an error of execution/commission? Was the error due to a “slip” (doing the right thing incorrectly) or a mistake (doing the wrong thing). Assume that it can’t be both.

Section 2 has two parts, identifying the types of medical error involved on the left side, and identifying which ACGME competency might be involved on the right side. For the medical error, there can be more than one. Put a “1” next to the most important error, followed by “2”, “3”, etc next to any additional errors you think are involved.

Section 3 asks you to identify in increments of 25%, how much of the complication that you scored at the beginning you would attribute to the error(s) you identified (the remainder being due to the patients disease processes rather than the treatment.

Thanks again for your help.

Appendix 5. Morbidity and Mortality Scenario 1 for External Validation

A 45 year old female was admitted to the hospital for an elective laparoscopic cholecystectomy. She had a prior hysterectomy through a Pfannenstiel incision a number of years ago and has had no other previous abdominal surgery. Prior to beginning the operation, we verified a lower abdominal “smile” incision just above the pubis. An infraumbilical curvilinear incision was made and the umbilicus mobilized. The umbilical ring was bluntly dilated with a Kelly clamp and a port was introduced into the peritoneal cavity. As soon as the scope was introduced, succus was seen. The procedure was converted to open and a tear was confirmed in a loop of ileum adherent to the umbilicus. This was repaired and the gall bladder was removed. There were no postoperative complications.

Errors: 1. Judgment 2. Technique

Appendix 6. Morbidity and Mortality Scenario 2 for External Validation

A 72 year old male was admitted through the Emergency Department after presenting with a complaint of left back pain. He is known to be hypertensive and a smoker. An abdominal CT scan confirms the presence of an abdominal aortic aneurysm with a confined area of leak into the retroperitoneum. After initial stabilization, he is brought to the operating room and undergoes a transabdominal aortic aneurysm repair with a tube graft. There is no back-bleeding noted from the inferior mesenteric artery orifice, which is oversewn prior to placing the graft. The iliac vessels are calcified but not aneurysmal. Prior to abdominal closure, the sigmoid colon is noted to be normal and signals can be heard on the antimesenteric border with the Doppler. Postoperatively he does well and is transferred from the intensive care unit on POD #2. On POD #5, he notes lower abdominal pain which progresses. Abdominal films are obtained which demonstrate free air. He is returned to the operating room and is found to have an ischemic perforation of the sigmoid colon. The sigmoid colon is resected and an end colostomy with closure of the distal colon (Hartmann procedure) are performed. The case is discussed at Morbidity and Mortality conference in detail and it is determined that the standard of care had been met, and with discussion of possible alternatives of postoperative care including a second look procedure and postoperative proctoscopy. On POD #5 after his colostomy, the patient has still not opened up. Abdominal films confirm gas through the GI tract, including a somewhat dilated colon. A gastrograffin enema per rectum demonstrates passage of the contrast into the colostomy bag. The patient is taken back to the operating room where it is noted that the distal rectum was brought up as the colostomy and the end of the sigmoid colon was left stapled and oversewn. This was corrected and the patient recovered without incident.

Error: 1. Inattention to detail

Appendix 7. Morbidity and Mortality Scenario 3 for External Validation

A 56 year old female was admitted through the emergency room with signs and symptoms of intestinal obstruction. She did not have any evidence of an acute abdomen, although her abdomen was tender to palpation. She was noted to have had multiple prior operations, and she had a large piece of mesh placed in her anterior abdominal wall to repair an incisional hernia at her last operation. She had received radiation therapy to the pelvis in the remote past. She weighs over 300 pounds and has had prior admissions for intestinal obstruction which have all resolved without need for operation. Abdominal films demonstrated scattered air-fluid levels at multiple levels of the small bowel and minimal gas in the colon, but with no free air or pneumatosis. A CT scan of the abdomen was performed which confirmed dilated loops of intestine and raised the question of a closed loop of bowel, although there was gas apparently distal to this area. Her WBC was 20,000, but the remainder of her laboratory studies, including amylase and blood gases, were normal. She was taken immediately to surgery. At operation, there was no clear transition point noted, no obvious closed loop obstruction, and no ischemic bowel. The mesh was densely adherent to the intestine and several enterotomies were made, necessitating resection of a segment of small intestine with primary anastomosis. Her abdomen was closed by reapproximating the prior mesh, which was densely incorporated into the abdominal wall without any separation. On postoperative day three, she was noted to have intestinal contents draining from her incision. Her fascia was opened locally and the site of drainage from the intestine could be seen in the open wound.

Error: Judgment

Appendix 8. Scenario Script: Operating Room

Scene opening: Two men, one older, one younger, in scrub suits at the scrub sink scrubbing their hands for surgery. The older is the surgeon and the younger is a resident.

Surgeon: Gee, Bob. I really wish I didn't have to do this complicated redo case now after being up all night with that trauma patient.

Resident: Why don't we ask the OR supervisor and the nurse anesthetist if we can delay the case until this afternoon so that you can get some rest. The patient isn't asleep yet.

Surgeon: No. That wouldn't be fair to the patient. Besides, I was trained to work without any sleep.

Fade out and fade into the middle of the surgical procedure. The surgeons are in the middle of the operation. The anesthetist is behind the ether screen, but visible. The scrub nurse is there, next to the surgeon. The surgeons appear to be working deep in the pelvis.

Nurse Anesthetist: Are you losing much blood down there?

Surgeon: Nope, everything is pretty dry.

Short time passes

Nurse Anesthetist: Are you sure you're not losing blood?

Surgeon: We're fine here, Suzie, really.

Short time passes. The nurse anesthetist hangs a unit of blood. Short time passes.

Surgeon: Suzie, I think something is wrong, the blood looks awfully dark. Is everything okay? (continuing to work in the pelvis)

Short time passes:

Nurse Anesthetist: We have a problem here. I've given a boatload of fluid and some blood. I'm having trouble with the blood pressure. I started some phenylephrine to bring his pressure up.

Surgeon: You gave him blood! Why didn't you tell me you were giving blood. We haven't lost much at all. You really ought to tell me before you give my patients blood. And when did you start the phenylephrine. Can you get the anesthesiologist in here right away?

Appendix 8. (Continued)

Anesthesiologist enters: Well, looking at the EKG I would say that we're in the middle of an MI. You better hurry up and get this patient to the ICU.

Fade out, fade in. Still in the OR.

Scrub nurse: It looks like we're missing a sponge, doctor.

Surgeon: That's all I need! (puts his arm into the pelvis) Nope, no sponges here. Let's close. I'm sure we have all the sponges out.

Scrub nurse: Don't you think we should get an x-ray?

Surgeon: How long is it going to take?

Scrub nurse: Probably ten minutes at most.

Surgeon: No, let's get out of here. This guy needs to get to the ICU.

Appendix 9. Armchair Conference

Set up like a Sunday morning talk show. One person in suit, three in scrub suits.

MC- Welcome, doctors. You had a chance to review the case. What are your thoughts?

Doctor 1. Well, I think there were several things that could have been improved. First, I think the surgeon has a responsibility to the patient to be awake and alert. Recent evidence is very clear that fatigue really does impair performance. The problem is that the surgeon isn't aware of the impairment. I think the surgeon should have made the effort to get the case postponed until the afternoon.

Doctor 2. I agree. But it isn't part of the culture of surgery. We all think that we can do anything and everything.

Doctor 3. It isn't part of the culture of the operating room either. But the surgeon is responsible for the patient and really needs to take a leadership role here.

Doctor 1. Then I think there was poor communication during the procedure. Generally when the anesthetist asks me if I'm losing blood, it means there is something wrong at the head of the table. As the surgeon, I need to take the initiative to ask a few questions. Maybe it wouldn't have made a difference in the outcome, but a half hour of lead time or a cardiologist in the OR wouldn't hurt.

Doctor 2. I agree. It isn't enough for the surgeon to be technically competent. The surgeon needs to be able to be aware of what's going on around him. In aviation, they call that situational awareness.

Doctor 3. I agree too. And this is more than the old "captain of the team" issue. The operating room is a complex place that requires frequent assessment of what is going on. The surgeon needs to be part of the solution.

Doctor 1. Finally, I think we have learned from experience that an incorrect sponge count is usually an incorrect sponge count. Most hospitals have protocols that require that a film is taken before the patient leaves the OR. I myself would have started closing the abdomen while they were getting x-ray. We would have had the film back before the skin was closed. If there were a sponge, and I'm pretty sure there was, we could have removed it while the patient was still in the OR.

Doctor 2. True. But I think it's more than that. A missing sponge is one of those problems that you can't talk your way out of. Surgeons really need to be more involved in paying attention to where sponges are left. It's one of those details that can avoid a

Appendix 9. (Continued)

lawsuit. Whenever I place a sponge in the abdomen, I make sure I tell the scrub nurse that there is a sponge in the abdomen. It doesn't take any time, and it helps me keep track.

Doctor 3. I can't disagree. I was sued once. Since then, keeping track of the sponges is one of those details I never overlook.

Doctor 1. So we agree that there was probably poor judgment in going ahead with the case.

Doctor 2. And we agree that the surgeon didn't try to understand the situation when the anesthetist was obviously trying to find a clue.

Doctor 3. And unless the patient is really crashing, and you're planning to come back, the surgeon needs to pay attention to details and follow the protocols.

Appendix 10. Follow-up Scenario Script: Without Errors

Follow-up script.

Surgeon and resident at scrub sink:

Surgeon: Gee, Bob. I really wish I didn't have to do this case now after being up all night with that trauma patient.

Resident: This redo could be really complicated. Why don't we ask the OR supervisor and the nurse anesthetist if we can delay the case until this afternoon so that you can get some rest. The patient isn't asleep yet.

Surgeon: You know, you may be right. I recently read that fatigue leads to a significant deterioration in ability to function, but that individuals are often not aware of the impairment. I would hate to harm this patient. I'll talk to the OR supervisor and nurse anesthetist, and if they agree I'll talk to the patient about waiting til this afternoon.

Fade out and fade into the middle of the surgical procedure. The surgeons are in the middle of the operation. The anesthetist is behind the ether screen, but visible. The scrub nurse is there, next to the surgeon. The surgeons appear to be working deep in the pelvis.

Nurse Anesthetist: Are you losing much blood down there?

Surgeon: Nope, everything is pretty dry. Is everything okay up there?

Short time passes

Nurse Anesthetist: Are you sure you're not losing blood?

Surgeon: Suzie, why don't I stop what I'm doing for a moment and let you figure out what's going on. (short pause) Is there anything that I can do to help you? (short pause) Perhaps we should get the anesthesiologist as well. Maybe three heads are better than one.

Nurse Anesthetist: Well there is no doubt that something is going on. Since you don't think it's due to blood loss, I guess we need to look into this further. Let's get Dr. Roberts (looking toward presumed location of circulating nurse).

Anesthesiologist: Well, well. What do we have here. I think we better look for some other options. Just glancing at the monitor, I wonder if we might have an evolving acute cardiac event. Let's get some cardiac enzymes. And let's look at the medications we've given. And let's start treating him for acute ischemia. And we better wrap up this operation as quickly as we can.

Appendix 10. (Continued)

Fade out, fade in. Still in the OR.

Scrub nurse: It looks like we're missing a sponge, doctor.

Surgeon: Can you quickly recheck the count. Since it looks like our patient may have had an MI, we really need to get to the intensive care unit as soon as we can. And you better call x-ray right away so that we don't lose any time. I can start closing while we're waiting for the film.

Scrub nurse: You told me to remind you that you put a sponge up over the liver. Did you take that one out?

Appendix 11. Fact Sheet - STAN, the METI Human Patient Simulator

Human Patient Simulator™ by Medical Education Technologies, Inc.™ (METI™)

METI's HPS-- Human Patient Simulator: Known as Stan (short for Stan D. Ardman or Standard Man) is a virtual patient capable of simulating nearly any possible human medical emergency including allergic reactions, the effects of nerve gas or weapons of mass destruction, drug overdoses, a severed limb or other traumatic injuries.

Creating a real life or death situation: Able to recognize and respond to medical treatment and drugs, once an emergency scenario is started, Stan becomes a real "life" placed in the hands of students that must diagnose and administer the correct treatment. Mistakes can send Stan into cardiac arrest and result in death, thus making any scenario, truly, life or death.

Lifelike human characteristics: Each Stan unit is built to simulate human characteristics and functions including blinking and dilating eyes, tearing, drooling, bleeding, pulsing, inhaling oxygen, exhaling carbon dioxide (with chest movement), talking, urinating, swelling tongue, etc.

Interactive Simulation: The Stan is completely interactive. Instructors use software to enter various emergency scenarios for which students are challenged to recognize according to the simulator's actions. The simulator then responds directly to the treatment as administered by the students. No interaction from the instructor is needed once the program has begun.

Students can practice skills over and over again until mastery is achieved.

Emergency Scenarios: METI has designed Stan to simulate physical (bleeding, pupil dilation) and physiological (pulse, heartbeat) characteristics for 100s of possible medical emergency situations.

Appendix 11. (Continued)

Who uses METI: Over 700 organizations worldwide utilize METI's technology including NASA, Center for Domestic Preparedness, U.S. and foreign military, leading medical schools such as Harvard, UCLA, Cleveland Clinic, Mount Sinai, Stanford and more.

Emergency Training Facts: The majority of emergency professionals enter the field having only practiced on lifeless mannequins and do not experience their first real emergency situation until after they have completed training and enter the workforce.

Training uses: Emergency preparedness, Disaster Training, Weapons of Mass Destruction, Bio Terrorism, Advanced Disaster Life Support, Advanced Cardiac Life Support, Advanced Trauma Life Support and more.

Appendix 12. Simulation Script: Recovery Room

Actors: PACU nurse, possibly anesthesiologist

Equipment: bed, patient, monitor (BP,HR), IV apparatus, chart, anesthesia record, preop record, EKG with acute MI from different patient, drugs (Beta blocker, levophed, epinephrine, oxygen,

Patient has been in the PACU for about an hour after a laparoscopic cholecystectomy. Unknown to the PACU nurse, the surgeon had bleeding from the cystic artery, which stopped spontaneously, and the cystic artery was never clipped. The surgeon and resident who did the procedure are scrubbed in the OR on another case. The PACU paged the resident doing the simulation, who has never met the patient and was not present during the operation, but is on the general surgery service.

Nurse: Thanks for coming doctor. This is Mister Bob Smith. He is 65 years old. He underwent a laparoscopic cholecystectomy today and has been here in the PACU for about 45 minutes. Dr. Miller did the operation, but he is scrubbed on another procedure in OR 3. The patient has been a little tachycardic since arrival and has received two 500 cc boluses of fluid. His blood pressure has been running between 90 and 100 the whole time and didn't seem to come up with fluid. The PACU anesthesiologist increased the maintenance IV fluids to 150 ml/hr and ordered an EKG. I gave Mister Smith 2 mg of IV morphine and his systolic fell to 70. We gave him another 500 cc of saline, but his pressure is still 70. He was complaining of some pain in his xyphoid area before the morphine.

About the Author

Peter J. Fabri M.D. is Professor of Surgery and Associate Dean for Graduate Medical Education and for VA Affairs at the University of South Florida, College of Medicine. Dr. Fabri obtained his Bachelor of Arts degree in chemistry from Northwestern University in 1969 and his Doctorate in Medicine from Loyola University-Stritch School of Medicine in 1973. He was elected to Alpha Omega Alpha, Sigma Xi, and Alpha Pi Mu. Dr. Fabri completed postgraduate training in General Surgery, Surgical Oncology, and Gastroenterology. Dr. Fabri has served in many academic and administrative positions over 30 years, including leadership positions in the ACGME, the American College of Surgeons, the American Association of Medical Colleges, the American Medical Association, the Department of Veterans Affairs, the JCAHO, the State of Florida Graduate Medical Education Committee, and numerous national surgical organizations and societies. He has over 150 peer-reviewed publications.