2002

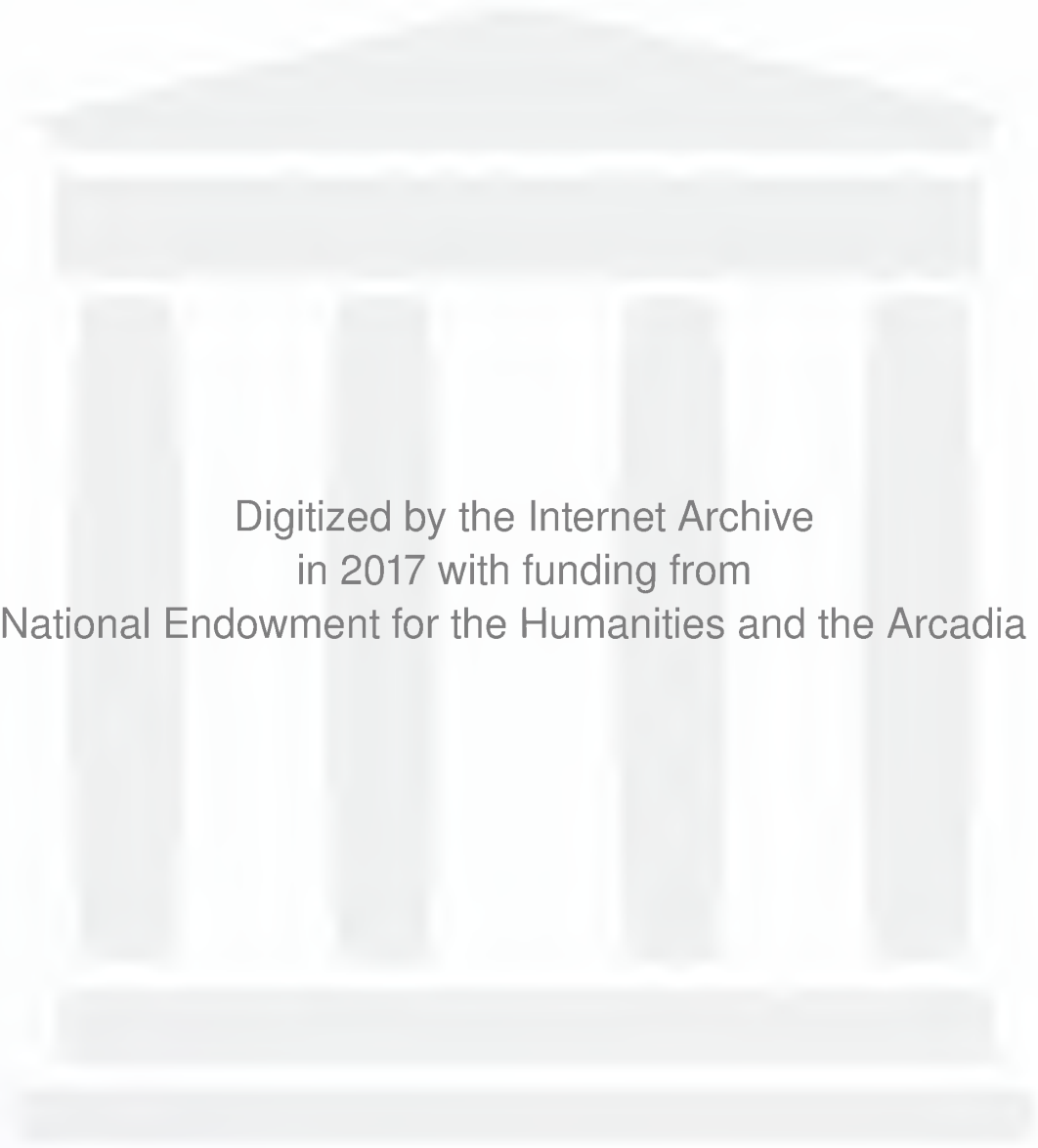
Evaluation of an Emergency Radiology Quality Assurance Program at a Level One Trauma Center and Clinical Consequences of Radiologic "Errors"'

Luke Sung-Wook Yoon
Yale University

EVALUATION OF AN EMERGENCY RADIOLOGY QUALITY ASSURANCE PROGRAM AT A LEVEL ONE TRAUMA CENTER AND CLINICAL CONSEQUENCES OF RADIOLOGIC "ERRORS"

Luke Sung-Wook Yoon

2002



Evaluation of an Emergency Radiology Quality Assurance Program at a Level One Trauma Center and Clinical Consequences of Radiologic "Errors"

A Thesis Submitted to the
Yale University School of Medicine
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

Luke Sung-Wook Yoon

ABSTRACT:

The purpose of this study was to evaluate the use of a redundant system in improving quality of care in the trauma setting by examining a subset of our Quality Assurance (QA) program. 531 consecutive abdominal/pelvic CT studies performed on trauma patients in a Level I trauma center from 08/22/99 to 08/21/00 were retrospectively reviewed. Each case was initially interpreted by a board-certified or board-eligible radiologist during the emergency department evaluation and was subsequently reviewed by a subspecialty abdominal imaging radiologist as part of a QA program. Nineteen were excluded due to incomplete information being available, resulting in 512 in our study. Cases with discordant interpretations had follow-up to discern management change. Of the 512 trauma cases, 153 cases showed discordant readings (29.9%). Review of patient records demonstrated changes in patient management in 12 cases (12/153; 7.8%). Three cases (3/153; 2.0%) were reviewed in morbidity and mortality records of the Department of Trauma Surgery as a direct result of misinterpretations. Six cases had additional diagnostic imaging studies for re-evaluation; 4/6 cases confirmed the QA reader's interpretation while 2/6 cases were shown to favor the initial interpretations. Our experience suggests that discordant radiologic interpretations most often do not result in a change in patient management and outcome. However, the QA program did identify and lead to changes in management of a number of cases by providing clinically significant additional findings.

ACKNOWLEDGEMENTS:

Many thanks to my thesis advisor Howard P. Forman, the Department of Radiology and the Department of Trauma Surgery.

TABLE OF CONTENTS:

INTRODUCTION:

Human error does occur in life. In a complex system such as medicine, human error is unavoidable. However, for people involved in medical care, the consequences of medical errors may be grave leading to serious injury or even death. We cannot simply accept that errors do occur; we must take action aggressively to study our health care delivery system, to identify areas of potential errors, and to redesign the system to prevent errors. Human error in medicine is considered as mismanagement of medical care induced by factors such as inadequacies in the design of a medical setting for the delivery of medical care, or cognitive errors of omission and commission precipitated by inadequate information or inappropriate mental processing of information (1).

Unlike other industries, such as aviation and military, human error in medicine has not been extensively researched and scrutinized for many years due to many reasons (2). First, medicine, as one of the most demanding professions, expects perfection from the providers, physicians in particular. The physicians have a difficult time admitting their mistakes and are not willing to learn from the errors. Second, the medical community did not foster a safe culture of reporting medical errors. In aviation, a safety culture is more than a set of guidelines; it is a behavior that governs the culture and belief of every member. With the existence of confidential incident reporting systems, pilots feel safe to report potentially disastrous incidents, and the industry in turn makes necessary changes for future prevention. On the other hand, in medicine, when it comes to errors, the focus has been on assigning blame to the person or the department associated with the error, rather than identifying the factors that contribute to the error.

Therefore, the topic of medical errors has not garnered much public attention despite some landmark studies published in the literature. For example, in 1991, the Harvard Medical Practice Study, a review of more than 30,000 charts from 51 New York hospitals, revealed adverse events in 3.7% of hospitalization (3). In 1994, Dr. Leape, one of the authors of the Harvard Medical Practice Study, called attention to the topic of error in medicine with the claim that 180,000 people die of iatrogenic injury each year (4). According to Leape, as many as 60% of these injuries were due to potentially preventable errors.

However, it was not until 1999 when the Institute of Medicine released the first report, To Error is Human: Building a Safer Health System, in a series of an Institute of Medicine initiative to develop a strategy for improving the quality of health care in America, the subject of medical errors became the focal point of public attention. The report sent shock waves throughout the medical community as it estimated that up to 98,000 Americans die each year as a result of preventable medical errors which was more than motor vehicle accidents, breast cancer, or AIDS (5). "Errors" were defined as "the failure to complete a planned action as intended or the use of a wrong plan to achieve an aim; not all errors result in harm." This report was careful not to assign blame on fallible caregivers but rather to expose the problem of our flawed health care system to prevent errors. This report called for an immediate action and recommended for 50 percent reduction in errors over the next five years.

Since the Institute of Medicine report in 1999, there have been numerous reports to expose, address, and recommend ways to reduce medical errors (6-9). The Quality Interagency Coordination Task Force responded to the Institute of Medicine report and to President Clinton by creating a center for patient safety within the Agency for Healthcare Research and Quality. Since then, this center in the Agency for Healthcare Research and Quality has conducted further research on medical errors and attempted to implement changes in our health care system as recommended by the Institute of Medicine. In 2001, the Institute of Medicine released a second report, Crossing the Quality Chasm: a New Health System for the 21st Century, to recommend a sweeping redesign of the American health care system and provide overarching principles for specific direction for policymakers, health care leaders, clinicians, regulators, and purchasers (6). Specifically, this report recommends that the Congress should create an "innovation fund" of $1 billion for use during the next three to five years to help subsidize promising projects and communicate the need for rapid and significant change throughout the health care system. One of the key areas relates to improvement of reporting systems and use of technological advances.

As we investigate the study of medical errors in the emergency department, we need to recognize that the emergency department is a unique place in the hospital. The previous Harvard Medical Practice study reported that 3% of adverse events occurred in the emergency department (10). Error in the emergency department differs from error in the rest of medicine (11). First, the nature of a typical emergency department necessitates intense time pressures to see patients and to triage them. Furthermore, inconsistent

arrival of patients makes the staff bored and less attentive during slow periods and harried during busy periods. In addition, most high-risk patients pass through an emergency department on the way into the hospital. Finally, the emergency department tends to be in flux where patients may be in any of many locations (in a room, hallway or radiology suite) and where staff rotate every shift. Thus, the study of medical errors in the emergency department requires an understanding that preventing error and ensuring patient safety in the emergency department involves different processes from other departments in the hospital.

The radiology department also faces unique challenges when dealing with medical errors since among the types of errors that may affect imaging patients are those due to misinterpretation. However, radiologic errors due to missed diagnoses are often difficult to ascertain as observer variation in interpretation does not necessarily represent medical error (12). Previous studies have investigated the subject of radiologic errors in general and of the frequency and clinical consequences of radiologic misinterpretations in the trauma setting (13-19). More specifically, two recent studies have examined occurrence and clinical consequences of radiologic errors in the emergency room. First, Wechsler et al. (13) compared the preliminary interpretation of emergency body CT scans by residents or fellows with the secondary review by attending radiologists and showed that major discordance occurs in 1.2% (7/597) and minor discordance occurs in 6.5% (39/597). In this study, there was no difference between discrepancy rates for trauma and nontrauma cases. Second, Eachempati et al. (14) sought to determine whether trauma patients could be discharged safely from the emergency department before the availability of official

readings for their radiologic examinations by evaluating alterations of preliminary readings in the emergency department and their effect on trauma patients. This study, like Wechsler et al., compared the preliminary interpretation by radiology residents with the secondary review by the attending radiologists by evaluating all radiologic studies performed in the emergency department in one year period. The result showed that only 102 of 38,260 discharged emergency department patients had official readings differing from preliminary readings. Of the 38,260 cases, 1073 cases were discharged trauma patients. Of the 102 cases that had discrepant preliminary and official readings, 42 were trauma cases. Thirty six of these 42 trauma cases were re-contacted for follow-up, requiring 8 repeat visits and 1 subsequent hospitalization. The study concluded that alterations of preliminary readings minimally affect outcomes of trauma patients. However, discharged trauma patients are more likely to harbor alterations of preliminary interpretations than other emergency department patients.

Other studies (15-19) also investigated the frequency and clinical consequences of radiologic errors in the emergency department. Lal et al. (15) evaluated the frequency of incorrect preliminary interpretations of neuroradiologic CT scans by on-call radiology residents and the effect of such misinterpretations on clinical management and patient outcome. This 9-month long prospective study compared preliminary interpretations by on-call radiology residents with second review by attending radiologists next day. The result showed that significant misinterpretations occurred in 0.9% (21/2388). There was a significant change in patient management in 12 of the cases, with a potentially serious change in patient outcome in two cases. Walsh-Kelly et al. (16) evaluated the clinical

impact of radiograph misinterpretation in pediatric emergency department and the effect of physician training level. Data were collected on 1,471 radiographs interpreted by pediatric emergency medicine attendings and emergency medicine, pediatric and family practice residents. These interpretations were then compared to the interpretation of a board-certified pediatric radiologist. The result showed 200/1471 (14%) misinterpretations. Non-radiology residents misinterpreted 16% of their radiographs versus 11% for pediatric emergency medicine attendings. Furthermore, only 20/1471 (1.4%) radiographs had clinically significant misinterpretations with no morbidity resulting from the delay in correct interpretation, demonstrating that radiograph misinterpretation by emergency department physicians occurs but is unlikely to result in significant morbidity.

In a different study, Roszler et al. (17) attempted to determine the accuracy of the residents' interpretations of emergency cranial CT scans done after working hours. During a 2-month period, a total of 289 cranial CT scans were retrospectively reviewed and the resident interpretation was judged acceptable, minor error, moderate error, or major error. The result showed that 6/289 (2%) neurologic examinations had four moderate and two major errors, with the mistakes all involving misinterpretation of cerebral hemorrhage. In another study done by Klein et al. (18), discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department was compared. In this prospective cohort study performed in a 13-month study period, 2083 radiographs were coded by the radiologist as concordant or discordant. Three hundred forty-nine of 2083 studies were coded as discordant. More importantly,

23/324 (7%), or 23/2083 (1.1%) overall, radiographs had potentially significant changes in patient management and outcome. This study concluded that the presence of radiologists to immediately read radiographs 24 hours a day could prevent missed findings. However, the cost effectiveness of such practice may not be justifiable given the small number of significant misinterpretations.

Lufkin et al. (19) had a different approach from the emergency medicine physicians' point of view with a hypothesis that radiologists' review of radiographs interpreted confidently by emergency physicians infrequently leads to changes in patient management. This prospective descriptive study compared radiologic interpretations between emergency department physicians and board-certified radiologists to determine whether radiologists' review is unwarranted when emergency department physicians are confident in their interpretations. The study showed that emergency department physicians were confident in 9,599 sets of radiographs out of a total of 16,410 (58%). Discordant interpretation rates for the "confident" and "not confident" groups were 1.2% and 3.1% respectively. Review of the 118 discordant interpretations in the confident group demonstrated that 11 were significant. Since total radiology review charges for the confident group were $215,338, the average radiology charge for each significant discordant interpretation was $19,576. This cost analysis, in the authors' opinion, did not seem to justify the standard practice of radiologists' review of all emergency department radiographs.

Overall, the literature review revealed only a few prospective studies of interventions designed to reduce reading error although several interventions, ranging from 24-hour radiologist review, to standardized checklists for high-risk misreadings, to regular conferences designed to prevent those errors may potentially show promises to reduce radiologic errors (11). Furthermore, the previously described studies (13-19) compared interpretations by radiology residents, fellows, or non-radiology attending physicians with attending radiologists, not between attending radiologists. Although these studies shed much light on occurrence and significance of discordant readings between physicians, the main objective of these studies was to determine the effect of training and experience in radiologic interpretations. Our study, in contrast, compares the interpretations between attending radiologists in order to investigate the rate and clinical significance of discordant interpretations and the use of redundant systems.

One of the characteristics of highly reliable industries includes high levels of redundancy in personnel and safety measures (5). For example, a Swiss chess model may be used to describe the redundant system as many layers in a system work to prevent error and maintain high quality. However, when the holes that appear in each layer happen to line up, an unfavorable error may occur. By creating more layers, one can prevent the chance that the holes in all layers line up at the same time. To achieve this, in April 1999, our institution established a new quality assurance (QA) system that complemented our existing 24 hour/day 7 day per week coverage by an attending radiologist in the emergency room. Every non-conventional radiographic imaging study done in an emergency department patient is interpreted by the attending radiologist in the emergency

department and subsequently reviewed by a subspecialty attending radiologist within the next 24 hours. During the first two years of this QA program, no formal analysis of the value of this approach has been undertaken.

In this study, we hypothesized that clinically significant improvement of patient management and outcome occurs with our quality assurance program. The purpose of our study, therefore, was to evaluate the use of a redundant system in improving quality of care in the trauma setting by examining a subset of our QA program. This study will serve not only as an internal review of the efficacy of the Yale radiology system, but it will provide valuable insight into reducing medical errors to prevent mortality and morbidity. By publishing our result in *Radiology*, we hope to contribute to the current ongoing research on the study of medical errors, particularly in emergency radiology.

STATEMENT OF PURPOSE AND HYPOTHESIS:

The purpose of this study is to evaluate the use of a redundant system in improving quality of care in the trauma setting by examining a small part of our QA program. We hypothesized that clinically significant improvement of patient management and outcome occurs with our quality assurance program. We sought to confirm this hypothesis by analyzing a subset of data for patient management and outcome by focusing on abdominal/pelvic CT studies performed in the setting of acute trauma. We conducted a retrospective study of abdominal/pelvic CT studies performed on trauma patients for one year and evaluated the data for the frequency and clinical consequences of misinterpretations.

METHODS:

We retrospectively reviewed 531 consecutive abdominal/pelvic CT studies performed on trauma patients, in an urban university-affiliated Level I trauma center, from 08/22/99 to 08/21/00. Nineteen studies which did not contain the QA reader's comments or names were excluded from analysis, resulting in 512 in our study. Further excluded are 182 chest/abdomen/pelvis CT studies, 11 pelvic CT studies, and 2 abdominal CT studies. Seven follow-up abdominal/pelvic CT studies performed on previously studied patients were also excluded from analysis. As mentioned previously, since April, 1999 in accordance with our QA program, every non-conventional radiographic study (CT, MR, ultrasound, and nuclear medicine) done in an emergency department patient has received a preliminary interpretation by an attending radiologist, "the primary reader", in the emergency department and a secondary review by a subspecialty attending radiologist, "the QA reader", within 24 hours of the initial interpretation.

The original report is generated during the emergency department evaluation by the primary reader using a voice recognition system, thus allowing for the immediate generation of a hard-copy text report. The QA report is generated through hand-written comments on a copy of the original report, with the QA reader's initials. The report is then returned to the primary reader for re-review. At the discretion of the primary reader, the report is addended. When there is a major discordance, the QA reader immediately contacts the primary reader as long as he or she is available. The case is discussed; and the clinicians are subsequently contacted. If the primary reader cannot be reached, the

QA reader contacts the clinicians immediately and may addend the report. All the QA reports are archived after the recheck process is complete. The Radiology Information System (IDX Rad, IDX Corporation, Burlington, VT) archives final reports but does not incorporate the rechecking physician or his comments in the electronic record. The 512 consecutive abdominal/pelvic CT studies in our study represent a subset of our overall data and include both adult patients and pediatric patients in the ED. This study was approved by the Human Investigation Committee of our institution. Informed consent was not required by the Human Investigation Committee for this study.

For each case, name, age, sex, clinical indication, names of primary and QA readers, traumatic abdominal/pelvic findings, traumatic extra abdominal/pelvic findings, and incidental findings were obtained and recorded. All 512 studies were then divided into two main categories: 1) complete concordance of interpretations and 2) discordance of interpretations. The findings identified by the QA reader and handwritten on the original report were considered "discordant." Discordant findings were then further categorized into three sub-categories: a) discordance of incidental, non-clinically significant findings, b) concordance of findings but discordance of interpretation, and c) discordance of potentially clinically significant findings. The categorization of the findings, when ambiguous, was determined by the consensus of three readers. Comments by the QA reader regarding anatomic variation (e.g., normal sized retroperitoneal nodes or retro-aortic left renal veins) or incidental observation (e.g., tampon in vagina; correctly placed nasogastric tube or Foley catheter) were not considered as discordant readings. The findings were then re-organized into three categories: 1) abdomen/pelvis trauma, 2)

non-abdomen/pelvis trauma, and 3) incidental findings. Data collected were stored and organized using a Microsoft Excel spreadsheet.

The primary readers consist of 21 board-certified or board-eligible radiologists with varying years of experiences ranging from less than one year to more than 20 years. Five out of the 21 primary readers were trained in body CT fellowship. One primary reader had a specific training in emergency radiology. The QA readers consist of 18 subspecialty radiologists also with varying years of experiences ranging from less than one year to more than 20 years. The QA reader reviewed all the cases regardless of the training background of the primary reader. For each individual reader (both primary and QA readers), a database using Microsoft Excel spreadsheet was created that shows the number of disagreed interpretations, the number of agreed interpretations, and the total number of cases read.

For the studies with discordant interpretations, additional data were obtained by 1) review of the patient medical record, 2) review of the correlated record of the Department of Trauma Surgery Morbidity and Mortality Conferences, and 3) re-evaluation of the final imaging reports and additional imaging studies. First, patient medical records of all the cases with discordant interpretations were obtained and reviewed to determine the clinical significance of these interpretations. Re-admission, new operation, new treatment, or new diagnostic studies (both imaging and laboratory) as a result of discordant second readings were considered clinically significant. Second, the records of the Department of Trauma Surgery Morbidity and Mortality Conferences were utilized to match "morbidity

and mortality" cases with discordant interpretations. The cases with a "positive match" were further reviewed to determine whether morbidity and/or mortality resulted from radiologic misinterpretation or other unrelated issues. Third, the final diagnostic imaging reports on all the cases with discordant interpretations were obtained, reviewed, and classified as "no change", "edited", or "with an addendum". For each case, subsequent imaging studies were reviewed by using IDX Rad, and new imaging studies as a result of discordant readings were used to determine whether the preliminary or the QA interpretation was accurate.

RESULTS:

Of the 512 trauma cases, 153 cases (153/512; 29.9%) showed discordant readings between the preliminary interpretation by attending radiologists in the ED (primary readers) and the QA review by subspecialty abdominal imaging radiologists (QA readers). The 512 studies comprise only abdominal/pelvic CT studies performed on the initial ED encounter with complete information on QA readers' identification and comments.

Review of all 153 patient records demonstrated that change in patient management occurred in 12 cases (12/153; 7.8%) (**Table 1**). One re-admission occurred as the patient was found to have adrenal hemorrhage by the QA reader. This patient, contacted at home, was subsequently sent home after physical examination and laboratory work showed no sequelae. In three patients, new diagnostic studies were requested and performed for suspected traumatic findings identified by the QA reader. In one patient, the QA reader identified possible pneumomediastinum, so the patient underwent swallow studies to rule out esophageal perforation. The result of the study was negative, and the patient was reassured. In the remaining two patients, the QA reader identified liver lacerations which were missed by the primary reader. Both patients were placed under strict bed rest, and serial hematocrit checks were performed for 1-2 days which delayed their discharge. Both patients were found to be stable and were safely discharged home.

In three patients, changes in patient management occurred although new findings identified by the QA reader were not trauma related (**Table 1**). Although the reason for ordering CT studies for these three patients was to rule out traumatic injuries, the QA

reader identified non-traumatic pathological findings in the CT studies that warranted further follow up. One patient received full laboratory evaluation (liver function tests, coagulation studies and hepatitis panel) for suspected cirrhosis that was identified for the first time by the QA reader. As a result of laboratory findings, the patient was diagnosed with hepatitis since hepatitis B and C antibodies were shown to be positive. The patient was also scheduled for an endoscopy for suspected esophageal varices as an outpatient. Another patient was found to have dilated loops of small bowels with thickened walls consistent with an inflammatory process by the QA reader. Gastroenterology was consulted, and the patient was treated with antibiotics. The last patient was found to have a left ovarian lesion suspicious for cystadenoma by the QA reader. The referring physician was notified of the finding. A new gynecology consult recommended follow up studies, but the patient refused further work-up in this case.

There are other changes in patient management that took place due to the new findings identified by the QA reader (**Table 1**). In one patient, the primary reader identified a mesenteric hematoma which turned out to be a normal variant as the QA reader reviewed the study. In this case, additional work-up for the patient was avoided due to the QA process. In three patients, new bone fractures were identified by the QA reader. In one patient, orthopedic and pain management was consulted for presumed acute vertebral compression fractures identified by the QA reader, and the patient had a corset placement to stabilize the fracture. The QA reader also identified a rib fracture in another patient who received Percocet for pain relief. In the third patient, there was a

questionable fracture at left ischium/pubic ramus transition. The patient was contacted, but the patient refused to come back to the hospital for re-evaluation.

TABLE 1

Clinical Consequences of Discordant Interpretation of Abdominal/Pelvic CT Scans on Trauma Patients

Patient age/ sex	Primary interpretation	Quality Assurance (QA) Interpretation	Change in management
8 months/ female	Mesenteric hematoma	No mesenteric hematoma	Avoided additional work-up
†44 years/ male	Normal	Liver nodules consistent with cirrhosis	New laboratory values ordered for LFT's, coagulation factors, hepatitis panel. Endoscopy appointment for varices as outpatient
17 years/ female	No fracture	Questionable fracture at left ischium/pubic ramus transition	Patient called but patient refused to come back to hospital
35 years/ female	No fracture	Transverse process fracture of L1	Orthopedic and pain management consults. Corset placement
33 years/ male	Normal	Right adrenal hemorrhage	Patient brought back to emergency room for re-evaluation and discharged after normal exam
50 years/ male	Normal	Pneumomediastinum	swallow studies to rule out esophageal perforation
*65 years/ male	Normal	Liver laceration	Strict bed rest. Checked hematocrit every 6 hours. Discharge delayed for 3 days.
*80 years/ female	Normal	Bladder rupture	Urology consult with Foley catheter placement
92 years/ male	Normal	Rib fracture	Percocet given for pain
*88 years/ female	Normal	Multiple liver lacerations	Trauma surgery consult. Checked hematocrit every 24 hours. Discharge delayed for 2 days
†31	Normal	Dilated loops of small bowels with	Gastroenterology consult.

years/ male		thickened walls consistent with an inflammatory process. May be due to Crohn's disease or an infectious process	Treatment with antibiotics
†91 years/ female	Normal	Left ovarian lesion suspicious for cystadenoma	Gynecology consult. Patient refused further workup

*cases also recorded in morbidity and mortality records in the Department of Trauma.
† non-trauma findings.

Abbreviations
LFT's : liver function tests

Our attempt to correlate the morbidity and mortality cases of the Department of Trauma Surgery with the 153 discordant cases resulted in 13 "matched" cases. Of the 13 "matched" cases, 10 cases were unrelated to diagnostic imaging studies. Three cases were directly related to delay in reporting of discordant diagnostic imaging findings (**Table 1**). Of these, two were due to delay in diagnosis of liver lacerations that required further laboratory evaluation to monitor hematocrit and resulted in lengthened hospitalization stay. One case involved a suspected bladder injury that required urology consult, placement of a Foley catheter and an additional imaging study for re-evaluation which showed normal bladder.

Review of the final diagnostic imaging reports in the discordant cases demonstrated 1) no change made to preliminary reports in 95 cases (95/153; 62.1%), 2) changes edited into final reports in 27 cases (27/153; 17.6%), and 3) addenda to final reports in 31 cases (31/153; 20.3%). Furthermore, review of subsequent diagnostic imaging studies for the discordant cases showed that 6 cases (6/153; 3.9%) had additional diagnostic imaging studies for re-evaluation. Although the QA reader recommended various follow-up imaging studies for re-evaluation in 13 cases, follow-up studies were performed in only 6 cases. These data, however, only include procedures performed at our institution. As shown in **Table 2**, 4 out of the 6 cases confirmed the findings by the QA reader: two liver lacerations, bowel loops instead of anomalous veins, and a rib fracture with a hemorrhagic renal cyst. The remaining 2 cases favor the initial interpretation. In both cases, the suspected bladder injuries by the QA reader actually turned out to be normal by follow-up CT cystogram.

TABLE 2

Follow-up Diagnostic Imaging Studies in Cases with Discordant Interpretation

Initial interpretation	Second interpretation	New study ordered	Final interpretation	Consensus with QA
Normal	Liver lacerations	CT Abdomen/pelvis	Liver laceration	Yes
Normal	Liver lacerations	CT Abdomen/pelvis	Liver lacerations	Yes
Anomalous vein	No anomalous vein, bowel loops	CT Abdomen/pelvis	No anomalous vein. Normal bowel loops	Yes
Normal	Bladder rupture	CT cystogram	No bladder injury	No
Normal	Bladder rupture	CT cystogram	Diverticula in bladder. no bladder rupture	No
Normal	Rib fractures and left renal hypoattenuation	CT Abdomen/pelvis and ultrasound	Rib fractures and left hemorrhagic renal cyst	Yes

Lastly, **Table 3** classifies every radiologic finding made by the primary reader and the QA reader into various categories. All in all, there are 1,133 findings identified in 512 CT studies. 203/1,133 (17.9%) describe abdominal/pelvic trauma. 244/1,133 (21.5%) describe traumatic findings that occurred outside of abdomen and pelvis. 686/1,133 (60.5%) describe incidental findings as agreed by the consensus of three investigators. These findings were also categorized according to the criteria described in Methods: 1) complete concordance, 2) discordance of incidental, non-clinically significant findings, 3) concordance of findings but discordance of interpretation, and 4) discordance of potentially clinically significant findings. 892/1,133 (78.7%) shows complete concordance between the primary reader and the QA reader. 127/1,133 (11.2%) shows discordance of incidental, non-clinically significant findings. 29/1,133 (2.6%) shows concordance of findings with discordance of interpretation. 85/1,133 (7.5%) shows discordance of potentially clinically significant findings.

TABLE 3

Classification of Radiologic Findings

	abd/pelvis trauma	non-abd/pelvis trauma	incidental	TOTAL
Complete concordance	163	214	515	892
Discordance of incidental, non-clinically significant findings	0	0	127	127
Concordance of findings with discordance of interpretation	10	2	17	29
Discordance of potentially clinically significant findings	30	28	27	85
TOTAL	203	244	686	1133

DISCUSSION:

Previous studies on medical errors

Even before the Institute of Medicine report in 1999, there were studies that raised public attention regarding medical errors. In 1991, Brennan et al. published a landmark study on incidence of inpatient adverse events, including those due to negligence (3). This study was the first of the two studies that were based on the Harvard Medical Practice. The authors concluded that patients experience a substantial number of iatrogenic injuries and that more than a fourth of those are due to substandard care. In the same year, this group released another study that classified the adverse events as drug complications, wound infections and technical complications as the most common types of error (10). The result suggested that many errors are preventable and that the study of errors, epidemiology and prevention can reduce incidence. In 1994, Dr. Leape, one of the leading authors in the previous Harvard Medical Practice studies, proposed several reasons for high error rate in medicine compared to other industries (such as aviation) (4). One reason may be a lack of awareness of the severity of the problem in the medical community. Second, most errors in medicine do no harm. But the most important reason is that physicians and nurses have a great deal of difficulty in dealing with human error when it does occur. This stems from the expectation that providers function without error as role models in medical education reinforce the concept of infallibility. Finally, the realities of the malpractice threat provide strong incentives against disclosure or investigation of errors. Leape suggested that the first step in reducing medical error is for practitioners to accept that they are fallible. Then as contributing factors are recognized

and studied, adverse events can be anticipated and reduced. However, although these studies received some attention around the medical community and the media, the subject of medical errors did not become the focal point of public attention until 1999.

In 1999, the Institute of Medicine (IOM) released the first report, To Error is Human: Building a Safer Health System, in a series of an initiative to develop a strategy for improving the quality of health care in America. The report sent shock waves throughout the medical community as it estimated that up to 98,000 Americans die each year as a result of preventable medical errors which was more than motor vehicle accidents, breast cancer, or AIDS (5). "Errors" were defined as "the failure to complete a planned action as intended or the use of a wrong plan to achieve an aim; not all errors result in harm." This reported recommended for 50 percent reduction in errors over the next five years and provided a four-tiered approach to implement changes. First, it recommended establishing a national focus to create leadership, research, tools, and protocols to enhance the knowledge base about safety within the Agency for Healthcare Research and Quality (AHRQ). Second, the IOM report called for identifying and learning from medical errors through both mandatory and voluntary reporting systems and at the same time protecting reporting systems from being used in litigation. Third, the IOM report sought to raise standards and expectations for improvements in safety through the actions of oversight organizations, group purchasers, and professional groups. Fourth, the IOM report recommended implementing safe practices at the delivery level.

Building upon the first report, the Institute of Medicine released the second report, Crossing the Quality Chasm: a New Health System for the 21st Century, on medical errors in 2001, outlining the major steps that should be taken in overhauling the U.S. health care system (6). This report suggested that Congress should create an "innovation fund" of $1 billion to help subsidize promising projects and communicate the need for rapid and significant change throughout the health system. Furthermore, this report detailed a 5-part strategy for building a stronger health care system. First, the report encouraged improvements in six areas in patient care to be safe, effective, patient-centered, timely, efficient, and equitable. Second, ten new rules to redesign and improve care in guiding patient-clinician relationships were introduced: care based on continuous healing relationships, customization based on patient needs and values, the patient as the source of control, shared knowledge and the free flow of information, evidence-based decision-making, safety as a system property, the need for transparency, anticipation of needs, continuous decrease in waste, and cooperation among clinicians. Third, health care system should be focused on the development of evidence-based approaches to care, especially in treatment of chronic diseases. Fourth, more supportive organizational process among health care organizations, clinicians and patients need to be created. This part of the five-part strategy calls for use of information technologies. Lastly, the committee emphasized changes in four key areas: more effective processes for the diffusion of clinical knowledge to providers and patients, use of information technologies to support clinical decision making, change in methods of payment, and appropriately preparing the work force for new challenges.

In response to these reports, two government groups, the Agency for Healthcare Research and Quality (AHRQ) and the Quality Interagency Coordination (QuIC) Taskforce, have taken action to implement changes in health care, inform the public, and provide research opportunities for studying medical errors (7,8,9). By February 2000, the Quality Interagency Coordination (QuIC) Task Force responded to the Institute of Medicine report and to President Clinton. In this report (7), the QuIC Task force listed each IOM recommendation from To Error is Human: Building a Safer Health System alongside responsive actions the QuIC will take in an errors reduction agenda with the creation of a center for patient safety within the Agency for Healthcare Research and Quality. Since then, this center in the Agency for Healthcare Research and Quality has conducted further research on medical errors and attempted to implement changes in our health care system as recommended by the Institute of Medicine.

One such area of research involves the identification and reduction of diagnostic errors and the study of system-specific causes (8). First, diagnostic inaccuracies may lead to incorrect and ineffective treatment or unnecessary testing, which is costly and sometimes invasive. For example, in obstetrics and gynecology, one study showed that physicians who performed 100 or more colposcopies a year had more accurate findings than physicians who performed the procedure less often (20). Likewise, in diagnostic imaging, studies that compared resident versus attending radiologists have shown that experience appeared to decrease discrepancy rates (13). This study investigated the effects of training and experience in interpretation of emergency body CT scans by comparing discrepancies between junior residents, senior residents and fellows with

attending radiologists. Of 598 CT studies, fellows demonstrated statistically significantly lower discrepancy rates than did senior or junior residents (5.9%, 13.7%, and 13.3% respectively). Second, although errors in medication, surgery, and diagnosis are the easiest to detect, medical errors may result more frequently from the organization of health care delivery and the way that resources are provided to the delivery system (8). The study of system-specific causes of medical errors is more difficult to perform and involves many more variables. Our study attempts to address these two issues, the identification and reduction of diagnostic errors and the study of system-specific causes, by studying the use of a redundant system to detect and correct image interpretation errors in the trauma setting.

Characteristics of highly reliable industries include an organizational commitment to safety, high levels of redundancy in personnel and safety measures, and a strong organizational culture for continuous learning and willingness to change (5). Use of redundant systems has been successfully employed in other industries such as military aircraft carriers or chemical processing. By providing multiple layers of check points, use of redundant systems in aviation has dramatically reduced potential disasters.

Aviation is an industry that depends its existence on safety. In aviation, a safety culture is more than a set of guidelines; it is a behavior that governs the culture and belief of every member. Helmreich in his work, Culture at Work in Aviation and Medicine (2), discusses and compares error management in aviation and medicine. In aviation, there exists a professional culture that actively encourages discussion, research, and strategies

to prevent potential errors. Helmreich points out that both in aviation and medicine there are five precepts for error management (2):

1. In any complex system, human error is inevitable. In systems such as aviation and medicine, where teams interact with technology, errors will occur.
2. There are limitations on human performance. All humans have limits imposed by cognitive capabilities such as the capacity of memory.
3. When performance limits are exceeded, humans make more errors. When overloaded or under stress, decision-making ability is hampered.
4. Safety is a universal value. In every culture, members value and strive to increase it. Safety does not come free although organizations differ in the resources they can devote to safety.
5. High-risk organizations have a responsibility to develop and maintain a safety culture. The task is to make sure that individuals and teams accept their responsibility for safety and error management.

Although there are many commonalties between aviation and medicine, aviation appears to be far ahead in reduction and management of errors. To achieve the highest level of safety, the airline industry aggressively pursues the use of redundant systems to provide multiple layers to check points to prevent errors. Furthermore, it devotes a lot of resources to conduct research to study, learn, and improve the existing system. Finally, in aviation, incident reporting systems are strictly confidential in order to promote a safe environment for learning from potential errors rather than a hostile setting that assigns blames on the individuals involved.

Errors in radiology and evaluation of our department's QA process

Using the concept of highly redundant systems to improve safety, in April 1999, our institution established a new quality assurance (QA) system that complemented our existing 24 hour/day 7 day per week coverage by an attending radiologist in the emergency room. As described previously, every non-conventional radiographic study (CT, MR, ultrasound, and nuclear medicine) done in an emergency department patient has received a preliminary interpretation by an attending radiologist, "the primary reader", in the emergency department and a secondary review by a subspecialty attending radiologist, "the QA reader", within 24 hours of the initial interpretation. Our study examined the use of this redundant system in improving quality of care in the trauma setting.

As discussed in introduction, previous studies have investigated the subject of radiologic errors in general and of the frequency and clinical consequences of radiologic misinterpretations in the trauma setting (13-19). Although many of the errors are due to disagreement in interpretations and often do not result in a change in clinical management and outcome, some of the "missed" findings do result in unfavorable clinical consequences. Studies by Wechsler et al. (13) that compared the preliminary interpretation of emergency body CT scans by residents or fellows with the secondary review by attending radiologists and by Eachempati et al. (14) that evaluated alterations of preliminary readings in the emergency department and their effect on trauma patients compared the preliminary interpretation by radiology residents with the secondary review by the attending radiologists. Other studies (15- 19) also investigated the frequency and

clinical consequences of radiologic errors in the emergency department. All of these studies (13-19), however, compared interpretations by radiology residents, fellows, or non-radiology attending physicians with attending radiologists, not between attending radiologists. The main objective of these studies was to determine the effect of training and experience in radiologic interpretations.

Our study is different from previous investigations in that we compared the interpretations between attending radiologists, focusing on one subset of our QA program: abdominal/pelvic CT studies performed on trauma patients. Our discordance rate of 29.9% (153/512) is higher compared to the previous study by Wechsler et al. (13). However, there are important differences between these two studies. Our study retrospectively reviewed discordant interpretations between attending radiologists while Wechsler et al. prospectively examined discordant interpretations between residents or fellows and attending radiologists. Although 153 of 512 cases had discordant interpretations, only 12 of 153 cases resulted in perceived changes in patient management. One case was of major concern as the patient needed to return to the emergency department for re-examination. The other 11 of 12 cases required additional diagnostic studies, laboratory values, new medications for pain and possible infection, and specialty consults. It is also important to note that 3 of 12 cases are due to significant non-trauma findings: a suspected cirrhosis, an inflammatory small bowel process, and a suspected ovarian cystadenoma. In the remaining 141 of 153 cases, new findings made by the QA reader did not affect the clinical management of the patients.

The average error rate among radiologists has been around 30% from studies dating from 1949 to 1992 (21, 22). In 1949, in his presidential address at the thirty-fourth annual meeting of the Radiological Society of North America, Dr. Garland (21) stated that radiologists are far less than perfect when it comes to accurately reading and interpreting radiographs. Discordance in interpreting radiographs was measured by a study of the relative frequency with which a reader was inconsistent with other readers (*inter*-individual variation) and with himself on two separate readings of the same set of films (*intra*-individual variation). The degree of *inter*-individual variation was from 9 to 24 percent. The degree of *intra*-individual variation was from 3 to 31 percent, which was surprising since the same reader missed the findings of the same set of films on two separate readings. Overall, interpretations of chest radiographs "missed" the pathological findings completely nearly 20% of the time, while close to 50% involved significant disagreements about the radiographic findings. This early study showed that the interpretation of radiographs is subject to a certain degree of inherent error and encouraged radiologists to be involved in improving the methods of describing lesions accurately and rational evaluation of existing classifications.

In 1975, after twenty-five years later, Herman et al. (23) obtained similar results among a group of Harvard University radiologists. Each of 100 chest radiographs, rich in abnormal findings, were read by five experienced radiologists who disagreed on the interpretation of chest radiographs as much as 56% of the time. Moreover, forty-one percent of the reports contained potentially significant errors. Three years later, the same group of researchers published another study that attempted to improve performance by

multiple interpretations of chest radiographs (24). Like the previous study (23), this study also had 100 chest radiographs, randomly selected from a hospital population, initially interpreted by eight radiologists. By using a method of duplicate reading, named "pseudoarbitration", a third independent interpretation was used to resolve disagreement between pairs of readers. This method reduced errors 37% and increased correct interpretations 18%. This study demonstrated the advantage of using multiple interpretations to improve in accuracy. Other factors such as implications for patient care and additional costs were considered and discussed.

In a more recent study at the Yale University School of Medicine, Elmore (25) revealed a disturbing variability in the radiologists' diagnostic interpretations, clinical accuracies, and management recommendations in reading mammograms. Radiologists in this study had substantial clinical disagreements in their diagnoses in up to 33% of the patients, and they disagreed in their management recommendations in up to 25% of the patients. The reasons for discordance according to the participating radiologists include differences in visual perception, differences in diagnostic criteria, and varying thresholds of concern. The researchers, led by principal investigator Alvan R. Feinstein, concluded that although mammography is of value in screening women for breast cancer, radiologists can differ, sometimes substantially, in their interpretations of mammograms and in their recommendations for management. Therefore, more efforts to improve accuracy and reduce variability in interpretation are needed to increase the effectiveness of mammography in detecting early breast cancers.

Although many of the errors are due to disagreement in interpretations and often do not result in a change in clinical management and outcome, some of the "missed" findings do result in unfavorable clinical consequences. Such "missed" findings often have far reaching ethical and legal consequences, and the ethical and medicolegal considerations of radiologic errors have been the subject of an ongoing debate for many years (26, 27). Leonard Berlin, who extensively studied the medicolegal issues in radiology and authored many articles in the topic, encouraged radiological societies on both national and local level to develop a "standard of radiological practice" which can be used for medicolegal purposes (27). Since errors in diagnostic radiology will continue to occur, we need to ascertain whether the error is due to negligence or not. If the error is due to negligence, which means that in the eyes of the court or jury no reasonable radiologist in similar circumstances would have made the error, then the defendant is guilty of malpractice and compensation to the injured patient is allowable. All interested parties should also provide review panels that would evaluate an alleged error and render an opinion as to whether or not it conformed to those standards. In his opinion, if such formal standards and review panels were developed and used successfully, the number of malpractice suits involving radiologists would decrease significantly.

In our study, it should be noted that the two readers, in each case, do not necessarily differ in their training level, as sub-specialty abdominal imagers function, at times, as primary emergency department radiologists. The difference, then, has much to do with the setting of the reading, and the proximity to clinical information. On the one hand, the emergency department radiologist is advantaged by knowing much more detail

about the current status of the patient, mechanism of injury and key clinical concerns. On the other hand, the environment of interpretation, in a level I trauma center, with very high volume, makes this setting less than optimal for the most diligent radiologists.

The other important area of finding involved the follow-up of the discordant cases. Of the 153 cases, only 58 cases (37.9%) showed changes (edits, addenda) made in the final reports. The remaining 95 cases (95/153; 62.1%) had identical preliminary and final reports. This finding suggests that the primary reader, more often than not, finds the QA reader's suggestion to be not significant enough to warrant changes to the original report. Further, review of subsequent diagnostic imaging studies for re-evaluation in 6 cases allowed us to ascertain whether the consensus lies with the primary reader or the QA reader. These additional imaging studies were performed at the QA reader's recommendation if the original studies raised any suspicion for pathologic findings which could not be adequately identified initially. In 4 of 6 cases, the subsequent studies agreed with the QA reader's interpretation. The remaining 2 cases favored the initial interpretation.

Our findings bring to light two important issues. First, our QA program serves an important purpose in identifying clinically significant, however infrequent, findings that are missed by the primary reader. The demonstration of changes in patient management suggests that the communication line between the QA reader, the primary reader, and the responsible clinician functions to improve patient care when needed. Second, despite the high rate of discordant interpretations (29.9%), most are not significant and do not result

in a change in patient management. Only 7.8% of the discordant readings (12/153) and 2.3% (12/512) of the total resulted in a change in patient management.

There are several limitations of this study. First, compliance with our QA system among the primary and QA readers is not perfect. There were 19 cases that were excluded from our data set because there was no name or comment from the QA reader. In our review of these records, there is no indication that these cases would represent "errors" as no follow-up imaging has occurred and no mention in the records of the Trauma Surgery department. Still, we cannot confirm what the QA findings would have been at the time. Second, this study did not provide the rate of accuracy of interpretations as measured against an infallible standard. Although the QA reader, with specialty training in body imaging, is often more experienced in reading body CT scans than the primary reader and certainly operating in a better setting for interpretation, in at least 2 cases the final interpretation favored the primary reader. Third, since this study focused on trauma patients in the emergency room setting, many of the recommendations made by the QA reader for further studies were often not followed up. After the patient is discharged, it is often difficult to contact and bring the patient back for further studies (14). Fourth, use of the medical record to identify cases that resulted in changes in clinical management may have been potentially biased by the reviewer's subjectivity.

In order to streamline our QA process, our department has recently hired a QA coordinator to oversee our QA program as well as to ensure that the process includes all cases, with appropriate documentation of the QA reader's findings and name. Other

efforts include encouraging more strict compliance by the attending radiologists and reducing the lag time between the initial interpretation and the QA review. Further, it is expected that the findings of such a program will eventually include proposals for remediation or continuing medical education, if a given primary reader is found to be deficient in an area of required expertise.

The cost of our QA program is relatively modest. For the most part, the attending radiologists on the Body-CT service spend one to two hours daily reviewing the previous 24 hour's cases. It is our estimation (after a sampling of 5 QA readers) that this process requires a full-time equivalent (FTE) for every 36 cases, and thus 14 days of a FTE are required for the total sample in this study. At our marginal cost of $800 dollars per day, this amounts to $11,200 for the detection of the 13 management-changing cases. Thus, the cost of detection is below $1000 per case. This is not the entire cost of the program as there are administrative costs and clerical labor, but this is a fair approximation of the marginal cost of professional time.

Another concern regarding the QA program pertains to liability. Although 153 of 512 cases contained discordant interpretations, it is presumptuous to label them as 153 "errors". Many of the 153 cases are often due to incidental additional findings of minimal clinical consequence. Reporting the "missed" radiologic diagnosis involves serious medicolegal and ethical considerations. Although our QA program is streamlined to report potentially significant missed findings and make necessary changes in the final report immediately, this practice is certainly not in place for many other institutions.

Rather, many radiologists in this country often face a dilemma when it comes to reporting "missed" radiologic diagnosis. For example (26), the following dilemma is familiar to many radiologists. The radiologist notices a spiculated, solid lung lesion with the typical appearance of carcinoma. He then checks the interpretation of the radiographs obtained a year ago and notices that the study was interpreted as normal. The radiologist then places the actual radiographs obtained one year ago on the view box and observes. To his dismay, the lung lesion was present on the original radiographs but was not noticed and thus not reported. A dilemma emerges: should the radiologists include in the report and inform the referring physician that the currently detected lesion was indeed present on previous radiographs but was missed, or should the radiologists remain silent on the content of the original images?

This dilemma occurs too often for radiologists. With the presumed discordance rate of 30% amongst radiologists, as discussed previously, the socioeconomic impact of "missed" diagnosis can lead to the growth in malpractice litigation, financial awards paid, and many aggrieved patients. Berlin (26) suggests that no single compelling argument can resolve this dilemma. However, the preponderance of legal opinion favors complete disclosure by the physician of all facts and information relevant to a patient's health or well-being, including complications of medical procedures and iatrogenic errors and injuries. Furthermore, from an ethical point of view, failure to disclose errors and mistakes constitutes an unethical conduct. For the radiologists reporting previously missed findings, they need to be careful when describing their findings, and words such as "missed", "error", or "mistake" should be avoided in official reports. To maximize

legal defense strategies for potential malpractice suits, the report of the misdiagnosis should be "succinct, matter-of-fact, and nonjudgmental" (26).

Conclusion: the impact of our findings

It is important that radiologists be interested in outcomes research. Outcomes research was initially defined in the United States Omnibus Reconciliation Act of 1986 as "research with respect to patient outcomes of selected medical treatments and surgical procedures for the purpose of assessing their appropriateness, necessity, and effectiveness" (28). John Thornbury, a renowned radiologist involved in outcomes research, encouraged the radiologists to be more involved in this area of outcomes research. In presenting the Eugene W. Caldwell Lecture at the annual meeting of the American Roentgen Ray Society in 1993, Dr. Thornbury clearly expressed his strong opinion (29) that if radiologists grasp the global outcome-oriented primary goal and become more involved and knowledgeable outcome-oriented consultants, they may then be influential in changing physicians' practices with regard to imaging selection and use. This will provide higher quality patient care and result in improvement of patient well-being. In this way, imaging examinations and interpretations will be optimally used for the most effective, efficient and highest quality patient care possible.

In order to assess the impact of this research on today's practice of clinical radiology and the patient management process in particular, we need to consider a hierarchical model of efficacy by Fryback and Thornbury (30). Efficacy is defined as "the probability of benefit to individuals in a defined population from a medical technology

for a given medical problem under 'ideal' conditions of use" (31). This hierarchical model of efficacy is presented as an organizing structure for appraisal of the literature on efficacy of imaging (30):

Level 1. Technical efficacy
- Resolution of line pairs
- Modulation transfer function change
- Gray-scale range
- Amount of mottle
- Sharpness

Level 2. Diagnostic accuracy efficacy
- Yield of abnormal or normal diagnoses in a case series
- Diagnostic accuracy (percentage correct diagnoses in case series)
- Predictive value of positive or negative examination (in a case series)
- Sensitivity and specificity in a defined clinical problem setting
- Measures of ROC curve height (d') or area under the curve Az

Level 3. Diagnostic thinking efficacy
- Number (percentage) of cases in a series in which image judged "helpful" to making the diagnosis
- Entropy change in differential diagnosis probability distribution
- Difference in clinicians' subjectively estimated diagnosis probabilities pre- to post-test information
- Empirical subjective log-likelihood ration for test positive and negative in a case series

Level 4. Therapeutic efficacy
- Number (percentage) of times image judged helpful in planning management of the patient in a case series
- Percentage of times medical procedure avoided due to image information
- Number or percentage of times therapy planned pretest changed after the image information was obtained (retrospectively inferred from clinical records)
- Number or percentage of times clinicians' prospectively stated therapeutic choices changed after test information

Level 5. Patient outcome efficacy
- Percentage of patients improved with test compared with without test
- Morbidity (or procedures) avoided after having image information
- Change in quality-adjusted life expectancy
- Expected value of test information in quality-adjusted life years (QALYs)
- Cost per QALY saved with image information

Level 6. Societal efficacy

- Benefit-cost analysis from societal viewpoint
- Cost-effectiveness analysis from societal viewpoint

According Thornbury, demonstration of efficacy at each lower level in this hierarchy is logically necessary, but not sufficient, to assure efficacy at higher levels. Applying this model, we can then assess the impact of our research on the field of radiology and today's health care system in general.

Our research, when compared to the Thornbury hierarchical model of efficacy, meets the criteria for levels 2, 3, and 4. Our study does not concern technical efficacy, so it does not meet the criteria for Level 1. However, according to Thornbury (32), Levels 2, 3, and 4 make up "clinical efficacy" for which our study meets all the criteria. Our study concerns diagnostic-accuracy efficacy (Level 2). By providing a second attending level radiologist as the QA reader, we are able to compare two attending level radiologists' interpretations to arrive at more accurate diagnosis. In 6 cases, when there were ambiguous interpretations, new imaging studies were performed and interpreted by another radiologist to determine consensus. Our study also affects diagnostic-thinking efficacy (Level 3). The communication between the primary radiologist and the QA radiologist often clarifies discordant interpretations and leads to change in diagnostic thinking process. Furthermore, the line of communication reaches further to the referring physician, who is then re-educated on new findings. Thus, the referring physician's diagnostic thinking is improved by our QA process. Our study demonstrated that our QA system affects therapeutic efficacy (Level 4). The review of the patient records and the

mortality and morbidity conference records in the Department of Trauma Surgery demonstrated that changes in patient management occur with our QA program. In this level, the patient participates with the physician in evaluating imaging results and making decisions about treatment choices. Some patients, when contacted about new findings, chose to come back to the hospital for further examinations while others refused. Finally, it is difficult to assess whether our study meets the criteria for Levels 5 and 6. A study that involves patient-outcome efficacy (Level 5) traditionally requires a prospective, randomized, controlled trial (32). At the highest level, societal efficacy (Level 6), the study design must be efficacious to the extent that it advocates changes in societal resources to provide medical benefits to society. Our QA program, despite its usefulness to our department, has not been proven to demonstrate and meet this highest level of efficacy.

Our findings show that clinically significant improvement of patient management does occur with a quality assurance program using redundant systems. Although most discordant interpretations do not result in a change in patient management, there are a number of cases in which patients are managed differently as a result of new clinically significant findings. As identification and reduction of medical errors become increasingly important in health care, evaluation of the existing quality assurance program, such as ours, will serve a useful purpose to monitor the efficacy of the current system and to make necessary changes to improve the system. Moreover, we believe that it will provide an invaluable educational experience for the housestaff and the attending radiologists as they learn from discordant interpretations as well as actual errors.

Through mutual feedback, both the primary and QA readers can improve areas of their weaknesses and instruct the residents on commonly missed findings.

Our QA system may serve as a model for utilizing the concept of redundant systems to prevent potential radiologic errors. We are not attempting to convince all other hospitals to adopt this QA program as it will be unrealistic for smaller hospitals with limited manpower in their radiology department. However, for larger academic medical centers with a medical school affiliation, we hope that they will take interest in our QA program and even consider adopting our program to suit their need. Our experience with the current QA system for the last three years has shown better coordinated care for the emergency department patients. Furthermore, emergency physicians and trauma surgeons developed deeper appreciation and trust in radiologists' interpretations. We, therefore, plan to continue with our current QA program for the foreseeable future. For now, there is no active discussion to expand our program to cover all studies performed at Yale-New Haven Hospital. We believe that current use of our QA program to cover the emergency department is sufficient to meet our pressing need without over-utilizing our resources.

Another study is currently underway to ascertain improvement of the accuracy rate due to our QA program by comparing the data before and after the institution of our QA program. This study will help us to have assurance that our QA system does indeed decrease the error rate. Moreover, although our brief cost analysis showed that the cost of our QA program is relative modest with about one third of a full time equivalent in the

entire department, we need more rigorous cost analysis to further improve our QA system to be more cost effective. Finally, we hope that our QA program will reduce discordance over time although we cannot predict that eventually there will be a time when the discordant rate will be low enough so that the system of reviewing the studies will not be justified. As we publish our study in *Radiology*, we sincerely expect that other academic institutions with adequate resources will consider our model to improve their radiology QA system as we strive toward our ultimate goal: reduction and prevention of radiologic errors.

REFERENCES:

1. Bogner MS. Human Error in Medicine. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1994.

2. Helmreich R, Merritt A. Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing Co., 1998.

3. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. *New England Journal of Medicine* 1991; 324:370-6.

4. Leape LL. Error in medicine. *JAMA* 1994; 272(23):1851-7.

5. Institute of Medicine. To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press, 1999.

6. Institute of Medicine. Crossing the Quality Chasm: a New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.

7. Quality Interagency Coordination (QuIC). Doing What Counts for Patient Safety: Federal Actions to Reduce Medical Errors and Their Impact. Washington, DC:

Report of the Quality Interagency Coordination (QuIC) Task Force to the President, Feb 2000.

8. *Reducing Errors in Health Care.* Translating Research into Practice, April 2000. AHRQ Publication No. 00-PO58. Agency for Healthcare Research and Quality, Rockville, MD, http://www.ahrq.gov/research/errors.htm.

9. *Medical Errors: The Scope of the Problem.* Fact sheet, Publication No. AHRQ 00 P037. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/errback.htm.

10. Leape LL, Brennan TA, Laird NM, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt HH. The nature of adverse events in hospitalized patients: Results of the Harvard Medical Practice Study II. *New England Journal of Medicine* 1991; 324(6):377-84.

11. Schenkel S. Promoting patient safety and preventing medical error in emergency departments. *Academic Emergency Medicine* 2000; 7:1204-22.

12. Robinson PJ. Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *British Journal of Radiology* 1997;70(839):1085-98.

13. Wechsler RJ, Spettell CM, Kurtz AB, Lev-Toaff AS, Halpern EJ, Nazarian LN, Feld RI, Needleman L, Alexander AA. Effects of training and experience in interpretation of emergency body CT scans. *Radiology* 1996;199(3):717-20.

14. Eachempati SR, Flomenbaum N, Seifert C, Fischer E, Hydo LJ, Barie PS. Alterations of preliminary readings on radiographic examinations minimally affect outcomes of trauma patients discharged from the emergency department. *J Trauma* 2000;48(4):654-8.

15. Lal NR, Murray UM, Eldevik OP, Desmond JS. Clinical consequences of misinterpretations of neuroradiologic CT scans by on-call radiology residents. *Am J Neuroradiol* 2000;21:124-9.

16. Walsh-Kelly CM, Melzer-Lange MD, Hennes HM, Lye P, Hegenbarth M, Sty J, Starshak R. Clinical impact of radiograph misinterpretation in a pediatric ED and the effect of physician training level. *American Journal of Emergency Medicine* 1995;13(3):262-4.

17. Roszler MH, McCarroll KA, Rashid T, Donovan KI. Resident Interpretation of Emergency Computed Tomographic Scans. *Invest Radiol* 1991;26(4):374-6.

18. Klein EJ, Koenig M, Diekema DS, Winters W. Discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department. *Pediatric Emergency Care* 1999;15(4):245-8.

19. Lufkin KC, Smith SW, Matticks CA, Brunette DD. Radiologists' review of radiographs interpreted confidently by emergency physicians infrequently leads to changes in patient management. *Annals of Emer Med* 1998;31(2):202-7.

20. Gordon P. Diagnostic accuracy of community physicians performing colposcopy. AHCPR Grant HS07162 Final Report; 1996.

21. Garland LH. On the scientific evaluation of diagnostic procedures: presidential address—thirty-fourth annual meeting of the Radiological Society of North America. *Radiology*. 1949;52:309-328.

22. Renfrew DL, Franken EA, Berbaum KS, Weigelt FH, Abu-Yousef MM. Error in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology* 1992;183(1):145-50.

23. Herman PG, Gerson DE, Hessel SJ, Mayer BS, Watnick M, Blesser B, Ozonoff D. Disagreements in chest roentgen interpretation. *Chest* 1975;68:278-282.

24. Hessel SJ, Herman PG, Swensson RG. Improving performance by multiple interpretations of chest radiographs: effectiveness and cost. *Radiology* 1978;127:589-594.

25. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *New England Journal of Medicine* 1994; 331(22):1493-9.

26. Berlin L. Reporting the "missed" radiologic diagnosis: medicolegal and ethical considerations. *Radiology* 1994;192(1):183-7.

27. Berlin L. Does the "missed" radiographic diagnosis constitute malpractice? *Radiology* 1977;123(2):523-7.

28. Report to Congress. *Progress of research on outcomes of health care services and procedures: introduction and overview.* Rockville, MD: Agency for Health Care Policy and Research, May 1991:1, AHCPR Pub. No. 91-0004.

29. Thornbury JR. Why should radiologists be interested in technology assessment and outcomes research? *AJR* 1994;163:1027-1030.

30. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11(2):88-94.

31. Brook RH, Lohr KN. Efficacy, effectiveness, variations and quality: boundary-crossing research. *Med Care* 1985;23:710-722.

32. Thornbury JR. Clinical efficacy of diagnostic imaging: love it or leave it. *AJR* 1994;162:1-8.

HARVEY CUSHING/JOHN HAY WHITNEY MEDICAL LIBRARY

MANUSCRIPT THESES

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the Medical Library are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but passages must not be copied without permission of the authors, and without proper credit being given in subsequent written or published work.

This thesis by
has been used by the following person, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS **DATE**