Errors of Interpretation as Elicited by a Quality Audit of an Emergency Radiology Facility[1]

James T. Rhea, M.D., Majic S. Potsaid, M.D., and Salvatore A. DeLuca, M.D.

A process-oriented quality care audit was performed in a large metropolitan hospital emergency radiology facility with an annual volume of over 50,000 examinations. One aspect of the audit dealt with errors found among interpretations by radiology residents, the initial interpreters of x-ray studies. Misinterpretations were identified by staff radiologists, who checked all examinations and countersigned the reports. Error rates were correlated with duration of training and were separated as to significance and whether the errors were false-negative (omission) or false-positive (commission). The false-positive to false-negative ratio was 27:73% which is in agreement with previous studies. For all cases of errors, the significance of change in interpretation was high in 20%, moderate in 29% and low in 51%. The effect of inadequate clinical history on the rate and significance of interpretation errors was also determined. When clinical information was inadequate, the significance was high in 27%, moderate in 40% and low in 33%.

INDEX TERMS: Diagnostic radiology, observer performance • Radiology and radiologists

Radiology 132:277–280, August 1979

TWENTY to forty per cent of statements in diagnostic radiology reports have been found to contain significant or potentially significant errors (1–6). During a quality assessment audit of the emergency radiographic area at Massachusetts General Hospital (MGH), the staff radiologists indicated that they had changed the residents' initial reports in a significant or potentially significant way in 11% of these reports. An analysis of the errors was made to determine the influence of initial errors and the revised interpretation on the patients' care and to investigate the characteristics of the errors.

While there will always be a base line or inherent error rate due to the limitations of the human process of perception and interpretation (1–9), an understanding of the factors influencing errors should help to keep the rate as close to the base line as possible.

METHOD AND DEFINITIONS

In the emergency area, radiology residents are on the firing line and issue an initial report, consulting with the staff radiologist (who is available from 8 A.M. until 11 P.M.) whenever necessary. From 8 A.M. until 11 P.M., two residents are present, usually a first-year and a second-year resident, who consult freely between themselves. From 11 P.M. until 8 A.M., only a second- or third-year resident is present. All resident reports are typed, combined with the films, and interpreted for a second time by a staff radiologist within a few hours of the resident's initial interpretation. If the staff radiologist makes a change considered to be relevant to the patient's care, the staff radiologist promptly informs the clinician of the change.

Over a 20-day period, 326 changes in the residents' reports out of a total of approximately 3,300 reports were analyzed. All changes were classified as false-negative or false-positive errors. A false-negative error occurred when the resident omitted a finding or stated something was normal when the staff felt an abnormality was present. If the resident failed to include a significant diagnostic possibility, this was considered a false-negative error (lung density called granuloma when tumor was also a possibility). A false-positive error occurred when the resident stated something was abnormal when the staff felt no abnormality was present. If the resident overemphasized the significance of an abnormality, this was considered a false-positive error (lung density called tumor when only granuloma should have been mentioned).

The above errors were further classified as errors of observation or errors of interpretation. Errors of observation included omissions as well as clearly measurable factors such as disk-space narrowing. Errors of interpretation included understating or overstating the significance of an abnormality, as well as errors resulting from technical factors such as rotation causing apparent clouding of the ethmoid air cells or a shallow inspiration mimicking pulmonary edema.

The changes were further classified as primary, secondary, and tertiary. A primary error was a significant or potentially significant error for patient care. Primary errors were subdivided into those related to the acute clinical question and those not related to it. Secondary errors were those of no significance to patient care. Tertiary errors were those of typing or grammar.

Subsequently, the charts of patients whose reports

[1] From the Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114. Submitted for publication 20 Nov. 1978; revision requested 6 Feb. 1979; accepted 17 April 1979. shan

TABLE I: NUMBER OF ERRORS

Type of Error	False Negative	(i)	(ob)	False Positive	(i)	(ob)	Total
1° A	97	(16)	(81)	33	(20)	(13)	130
1° B	64	(5)	(59)	13	(7)	(6)	77
2°	49	(11)	(38)	16	(14)	(2)	65
	210	(32)	(178)	62	(41)	(21)	
3°							54
							326

(i) = Interpretation error.
(ob) = Observation error.
1° A = Primary error related to the acute clinical question.
1° B = Primary error not related to the acute clinical question.

TABLE II: ERROR BY TYPE OF EXAMINATION

Examination Type	1° Error	2° Error	3° Error	Total	Per Cent
Skull and face	8	5	9	23	14
Lumbar spine and pelvis	7	5	3	15	14
Cervical spine	1	4	2	7	
KUB	6	4	2	12	8
Extremities	16	6	11	33	22
Chest and ribs	10	26	23	59	40
IVP	1	0	0	1	<1
				150	

TABLE III: TYPE OF EXAMINATION AS A PERCENTAGE OF TOTAL VOLUME

Type of Examination	Percentage of Total Volume
Skull and face	12
Spine and pelvis	11
Abdomen and GI	10
Extremities	29
Chest	37
Genitourinary	1

contained the primary errors were reviewed to determine the effect on the patient's care of the change in interpretation, which interpretation was correct (resident or staff) given the clinical follow-up, the significance of the error, and whether the clinical information on the radiology requisition form was adequate.

RESULTS

TABLE I gives the numbers of errors by type. Of the total number of errors, 33% were in reports by second-year residents and 66% by first-year residents who were in the latter part of their year of training. The second- and third-year residents interpreted 68% of all examinations and the first-year residents interpreted 32%.

TABLE II gives the errors by type of examination based on 150 consecutive errors. TABLE III presents the type of examination as a per cent of the total volume of examinations done in the emergency area and is based on 26,000 consecutive examinations.

Of the number of cases of emergency ward errors, the final interpretation was normal in 39%. Insignificant findings occurred in an additional 29%. Significant abnormalities relative to the clinical questions were seen in 25%. Other significant abnormalities not related to the clinical question were seen in 7% of cases.

Of the 161 primary errors, charts were obtained on 71% for evaluation of the clinical follow-up to further determine the characteristics of the errors. The change in interpretation had a beneficial effect on patient care in 37% of cases and caused unnecessary workup in 2%. The effect of the change in interpretation could not be determined from the chart in 61%.

It would be expected that a second interpretation would introduce new errors as well as correct initial errors. It was found that the second interpretation of the staff was correct in 22% of the cases; the resident's initial interpretation was correct in 5% of the cases; and in the remaining 73% of cases it could not be determined from the chart whether the resident or staff interpretation was correct.

In each case, the clinical information provided at the time of the examination was evaluated in light of the chart evaluation. The clinical information was adequate in 90% of cases and inadequate in 10%.

Lastly, the significance of the change in interpretation was graded and was considered to be high in 20%, moderate in 29%, and low in 51%.

DISCUSSION

Inherent Error Rate

The previously reported error rate in an observer's interpretation of an image is about 30% (20–40%) and is confirmed by this study (1–6). It has been shown that a second reading will discover about one third of the initial errors (8, 9). Assuming about 30% initial errors and that the second reading by the staff would discover a third of these, then the observed frequency of change in our initial reports, 11%, would be very close to the expected 10% found by others.

It has been inferred that such an error rate is in part due to an inherent limitation of the human perceptual process (1–9). There may be other factors which influence the error

TABLE IV: EXAMINATIONS IN WHICH THE FIRST INTERPRETATION WAS CORRECT

Case	Initial Classification of Type of Error	Initial Reading	Second Reading	Clinical Follow-up
1.	1° FP	Fracture of arm	No fracture	Fracture present
2.	1° FP	Fracture of radius	No fracture	Fracture present
3.	1° FP	Stress fracture of foot	No fracture	Fracture present
4.	1° FN	Distended colon	Colonic obstruction	Enema: no obstruction
5.	1° FN	Soft-tissue swelling	Question tendon injury	Physical: no tendon injury
6.	1° FN	Chest normal	Question metastasis of humerus	Bone scan: negative
7.	1° FN	Shallow inspiration	Abnormal right hilus	Subsequent films normal

FP = False positive.
FN = False negative.

rate which are controllable. However, we agree that there will always be errors when humans perceive and interpret an image; *i.e.*, a baseline or inherent error rate. The striking similarity of the characteristics of the errors in this study and previous studies lends support to the idea of an inherent error rate. If controllable factors were responsible, a greater variation in error characteristics would be expected among different studies of errors.

One such error characteristic which seems surprisingly constant is the false-positive to false-negative ratio. In previous studies this ratio, for significant and potentially significant errors, has been reported as 22:78% (2) and 20:80% (3). In this study the ratio is 27:73%. The two previous reports involved chest examinations, while this study involves all examinations performed in the emergency ward area, only 37% of which are chest examinations.

Another observation of this study which supports the concept of an inherent limitation of the perceptual process is the surprisingly close correlation between the type of examination as a percentage of total errors (TABLE II) and the type of examination as a percentage of total volume (TABLE III). There seems to be about the same chance of making an error when interpreting a chest examination as when interpreting an abdominal or extremity examination. However, it is interesting to note that there is a relatively greater proportion of primary errors in extremity examinations and a relatively greater proportion of secondary errors in chest examinations. The reason for this may relate to a greater incidence of secondary findings on the chest examination.

It has previously been noted that the percentage of errors increases as the significance of the error decreases (3). This inverse relationship not only is seen in this study, but the incidence of errors correlates very closely with other studies. Previously, the significance of errors was graded by the probable effect on patient care. It was found that 20% would affect, 25% would potentially affect, and 55% would have no effect on patient care (3). In this study similar criteria were used to determine the significance of the change in interpretation (TABLE IV). High, moderate and low significance were based on the change having an important effect, probably having an effect, and having little effect on the care of the patient; these observed percentages were 20%, 29% and 51%, respectively.

Factors Influencing the Error Rate

While the above findings lend support to the concept of an inherent limitation of the perceptual process, there are factors to be considered to make the basic error rate as small as possible. In a study of coal workers' pneumoconiosis (7), factors influencing the error rate were film quality and physician familiarity with roentgenographic manifestations of the disease. In the quality assurance audit of the emergency area, film quality was thought to be satisfactory by both resident and staff in 96% of the cases from 8 A.M. to 4 P.M., 94% of the cases from 4 P.M. to midnight, and in 86% of the cases from 12 midnight to 8 A.M..

In prior studies it has been found that there is no consistent pattern relating errors and the duration of training beyond the first year of residency (3). In this study, it was found that the first year residents made a relatively greater number of errors than the second year residents on their initial interpretation in spite of the somewhat better film quality when the first-year residents were working. The types of cases seen when only a senior resident was present might have been different, but 89% of the volume of cases occurred when both the first-year and more senior residents were present. The residents were in the latter part of their particular years of training.

An additional factor influencing the error rate also demonstrated by this study is the adequacy of clinical information (9). During a previous quality audit, both resident and staff considered the clinical information inadequate in only 5% of cases. However in this study, in those cases with errors, clinical information was inadequate in 10% of the cases and the percentage of significant errors was much greater in those cases with inadequate clinical information. For all cases of errors, the significance of the change in interpretation was considered high in 20%, moderate in 29%, and low in 51%. In those cases of error in which the clinical information was inadequate, the significance was high in 27%, moderate in 40%, and low in 33%.

Effect of Second Interpretation on Patient Care

It was impossible to determine the effect on patient care of the second interpretation or to determine which interpretation was correct in the majority of charts reviewed. Many patients were treated, sent home, and did not return. In other instances, the second report added information that would not change the treatment or workup. In still other instances when the diagnosis was changed to normal, no follow-up was obtained.

A second interpretation of the examination had a beneficial effect 20 times as often as it caused unnecessary workup. There were no harmful effects of the second interpretation except for limited added workup. There were 7 instances in which the first interpretation was proven correct clinically and in 3 of these additional workup was performed (TABLE IV).

The 3 cases in which additional workup was done which might not have been done in the absence of the second interpretation were numbers 4, 6, and 7 in TABLE IV. The additional workup included a barium enema, bone scan, and subsequent chest radiograph.

It is of interest that all cases in which the staff incorrectly thought a primary false-positive error had been made involved subtle fractures. This may be a result of the resident examining the patient or talking with the clinician about the site of injury at the time of the initial interpretation or the better visual acuity of younger observers.

SUMMARY

Additional evidence is presented to support the concept of an inherent error rate due to the perceptual process in radiologic interpretation.

Controllable influencing factors include film quality and the experience of physicians. It has been reported that the error rate in controlled studies does not vary with the level of training or experience after the first year of radiology residency. In this study, the first-year residents in the eleventh month of training were observed to make relatively more errors than second year residents in their twenty-third month of training. An additional factor increasing the incidence and significance of errors is inadequacy of clinical information available to the radiologist.

A second interpretation of an examination is an effective means of reducing errors. There is a low incidence of new errors which are introduced by the second interpretation. In this study, the only adverse effect on the patient which occurred was limited additional workup.

James T. Rhea, M.D.
Department of Radiology
Massachusetts General Hospital
Boston, MA 02114

REFERENCES

1. Tuddenham WJ: Visual search, image organization, and reader error in roentgen diagnosis. Radiology **78:**694–704, May, 1962.
2. Herman PG, Gerson DE, Hessel SJ, et al: Disagreements in chest roentgen interpretation. Chest **68:**278–282, Sep 1975
3. Herman PG, Hessel SJ: Accuracy and its relationship to experience in the interpretation of chest radiographs. Invest Radiol **10:** 62–67, Jan–Feb 1975
4. Berlin L: Does the "missed" radiographic diagnosis constitute malpractice? Radiology **123:**523–527, May, 1977
5. Revesz G, Kundel HL: Psychophysical studies of detection errors in chest radiology. Radiology **123:**559–562, Jun 1977
6. Swensson RG, Hessel SJ, Herman PG: Omissions in radiology: faulty search or stringent reporting criteria? Radiology **123:**563–567, Jun 1977
7. Felson B, Morgan WKC, Bristol LJ, et al: Observations on the results of multiple readings of chest films in coal miners' pneumoconiosis. Radiology **109:**19–23, Oct 1973
8. Hessel SJ: Personal communication
9. Taber F: Visual perception of roentgen images. CRC Crit Rev Diagn Imaging **10:**187–202, Dec 1977