ORIGINAL ARTICLE

Chest radiographs in the emergency department: is the radiologist really necessary?

M E Gatt, G Spectre, O Paltiel, N Hiller, R Stalnikowicz

Postgrad Med J 2003;79:214–217

Background: The chest radiograph is considered one of the most complex imaging modalities to interpret. Several studies have evaluated radiograph interpretation in the emergency department, and considerable disagreement among clinical physicians and expert radiologists has been observed in the reading of chest films. The interpretation of chest radiographs by emergency department physicians was compared with senior radiologists in discharged patients, and misinterpretations assessed in relation to the physician's level of training.

Methods: Radiological descriptions of 509 chest radiographs of 507 patients, aged 16–98 years who were discharged from the emergency department, were prospectively reviewed. Missed findings were recorded with regard to the physician's level of training and experience. The effects of misinterpretations on discharge recommendations were also investigated. Statistical assessment was conducted using the χ^2 test. Interobserver agreement was also tested by the κ coefficient.

Results: The sensitivity for detecting different abnormalities in the radiographs ranged from 20% to 64.9% and specificity from 94.9% to 98.7%. Despite the low sensitivities found, there were relatively few clinical implications of the "missed" findings since they were either of a minor nature or appropriate follow up was prescribed. The overall interobserver reliability, assessed by the κ coefficient, was 0.40 (95% confidence interval 0.35 to 0.46). These findings did not change significantly by emergency department physician's level of training.

Conclusions: Emergency department physicians frequently miss specific radiographic abnormalities and there is considerable discrepancy between their interpretations and those of trained radiologists. These findings highlight the importance of routine evaluation of chest radiographs by a well trained radiologist and emphasise the need for improving interpretive skills among emergency department physicians.

See end of article for authors' affiliations

Correspondence to:
Dr Moshe E Gatt,
Department of Medicine,
Hadassah University
Hospital, Mount Scopus,
PO Box 24035, Jerusalem
91240, Israel;
Gatt@md.huji.ac.il

Submitted 12 July 2002
Accepted
25 November 2002

Patients seen in the emergency department often undergo radiological examinations for evaluation of their medical and surgical conditions. The treating physician in the emergency department does not always have the time or the opportunity to consult an on-call radiologist, and therefore has to rely on personal experience and basic skills. Discordance of radiograph interpretation in the emergency department with evaluations of radiologists is commonly reported in different studies as ranging from 0.3%–17%,^{1–4} with one study reporting up to 58% discordance by primary care physicians.⁵ However, a change in treatment based on such discordance is required in only 0.06%–3% of patients. Intermediate levels of interobserver variability have been reported.^{2,6–9}

Previous reports showed higher rates of misinterpretation of the chest radiograph compared with other radiographs,^{3,5,6} yet none specifically address the chest radiograph and its complexity in adult patients. Among patients who are hospitalised there are more opportunities to review the initial treating physician's interpretation of radiographs with that of a senior radiologist. Hospitalised patients are easier to locate, treat, and re-evaluate if a discrepancy is detected. Patients discharged home from the emergency department obviously pose a difficulty in correction of misinterpretations. We therefore chose to investigate the subset of patients who were discharged from the emergency department and not all chest radiographs done in the emergency department as previously reported. In these patients errors in radiograph interpretation can affect outcome and sometimes may prove fatal. We also reviewed the effects of misinterpretations on discharge recommendations for further treatment, investigation and follow up, and assessed these results in relation to the emergency department physician's training and experience.

METHODS

The chest radiographs of 507 patients (509 examinations) were prospectively collected during a four month period from February to May 2000. The study population consisted of patients aged 16 years and older treated in the emergency department at Hadassah University Hospital, Mount Scopus. The files of discharged patients from the emergency department were reviewed on a daily basis. In contrast to other studies,^{3–5,10} in order to increase the sensitivity results and minimise bias, emergency department physicians were requested to note whether they consulted with a radiologist before discharging the patient and what their own interpretation was before the consultation was given. There was no standard form or checklist used to record the radiological interpretations.

Data recorded with the radiological examination included the physician's level of training, the patient's age, gender, complaints, physical examination, radiograph interpretation by the emergency department physician and senior radiologist, final diagnosis at discharge, and further recommendations for treatment and follow up. Chest radiographs for which there were no interpretations, or discharge letters in which the final diagnosis was not clear, were excluded. A senior radiologist's interpretation was considered the "gold standard" for the final interpretation. Misinterpretations occurred when the initial emergency department interpretation and that of the final senior radiologist were discrepant. Reviewing the

Abbreviations: CI, confidence interval; CHF, congestive heart failure

Table 1 Diagnoses of patients discharged from the emergency department who had a chest radiograph performed

Diagnosis	Frequency (%)
Non-specific*	230 (45.2)
URTI/bronchitis	79 (15.5)
Pneumonia	48 (9.4)
Trauma	47 (9.2)
Musculoskeletal disorders	42 (8.3)
Asthma/COPD	36 (7.1)
CHF	27 (5.3)

*Pleuritis, hyperventilation, no disease found, no definite diagnosis.
URTI, upper respiratory tract infection; COPD, chronic obstructive lung disease; CHF, congestive heart failure.

patient's medical records assessed the clinical significance of a misinterpreted radiograph.

Discrepancies of clinical significance were divided into three categories: mild clinical significance requiring no further evaluation (that is, pleural thickening, chronic changes, soft tissue and bone deformity, and old fractures); moderate clinical significance requiring further evaluation (that is, osteopenia, spinal fractures, cardiomegaly, hiatal hernia, interstitial markings, pleural and lung calcification, and parenchymal lesions); and high clinical significance requiring prompt evaluation and treatment (that is, consolidation, congestion, pleural effusion, mediastinal widening, new fracture, atelectasis, high diaphragm, and coin lesions). It should be noted that even subtle signs such as minimal changes in the chest radiograph were counted as misinterpreted if not specified by the emergency department physician.

The emergency department staff participating in the study included five board certified internal medicine attending physicians, five final year residents, five intermediate year (two to three years postgraduate) residents and seven first year residents, as well as six surgical residents at different levels of training. There were 10 board certified senior radiologists who evaluated the final radiograph diagnosis.

The accuracy of the emergency department compared with the senior radiologist's interpretations was expressed in terms of sensitivity and specificity. Differences in proportions were assessed using the χ^2 test. Interobserver agreement was also tested by κ coefficients and their 95% confidence intervals (CI) among the different levels of emergency department physicians.¹¹ For all statistical analyses a two tailed p value of 0.05 was considered statistically significant.

RESULTS

Of 507 patients 57.2% were male and 42.8% were female, their ages ranged between 16 and 98 years (mean (SD) 48 (20.5) years). The most frequent complaints were chest pain (23.9%), dyspnoea (18.7%), cough (19%), and fever (13.1%). Common physical findings were normal chest examination (36.8%), musculoskeletal tenderness (13.2%), inspiratory crackles and signs of bronchospasm (12% each). The diagnoses at discharge were mostly non-specific (such as "non-specific chest pain"), or diagnoses not related to radiography findings, as shown in table 1.

There were 557 findings described by emergency department physicians compared with 647 found by senior radiologists for all chest radiographs evaluated (each radiograph occasionally having more than one finding, and each finding was assessed separately). The emergency department physicians consulted a radiologist in 147/509 of the radiographs examined (28.9%). The major findings on radiographic interpretation and estimates of sensitivity and specificity are shown in table 2, and the overall sensitivity was relatively low. The highest level of sensitivity was shown for consolidation (64%) and congestion (50%), whereas very low levels of sensitivity were found for chronic changes (20%). Specificities were high, ranging from 94.9% for consolidation to 98.7% for pleural effusion. It is important to emphasise that the emergency department physicians often missed potentially important findings such as coin lesions or mediastinal widening, although the numbers were too small to allow precise statistical assessment.

We assessed the treatment actually received and compared it with the treatment "indicated" according to the radiologist's description. Twenty two of 57 patients (38.6%) with misinterpreted radiological signs of consolidation were discharged without a new antibiotic prescription. Of these, 10 were already receiving antibiotics or were known to have pneumonia on admission to the emergency department, so that the actual number of patients not receiving treatment for suspected pneumonia was 12 (26.3%). Three of them had been specifically instructed to return for further evaluation the next day, thus reducing the potential number of improperly treated patients to nine (15.8%). Twenty six of 34 patients (76.5%) with misinterpreted radiological signs of congestion were discharged without specific treatment or change in treatment for congestive heart failure (CHF) exacerbation. Twenty of these patients were known to have CHF and were receiving treatment, therefore reducing the actual number of patients not receiving appropriate treatment for CHF exacerbation to six (17.6%), of which three were sent for

Table 2 Sensitivity and specificity of emergency physicians' interpretations compared with those of senior radiologists in abnormalities found on chest radiographs

Category	Count by emergency department physicians	Count by senior radiologists	Sensitivity (%)	Specificity (%)
Consolidation	40	57	64.9	94.9
Congestion	19	34	50.0	97.5
Pleural effusion	9	31	25.8	98.7
Chronic changes*	8	37	20.0	97.0
Coin lesion	3	13	ND	ND
Mediastinal widening	2	8	ND	ND
New fracture	3	4	ND	ND
Deformations	6	22	ND	ND
Cardiomegaly	32	102	29.4	98.5
Non-specific changes†	20	48	39.5	96.4
Any chest radiograph abnormality	346	291	60.1	89.0

*Pleural thickening, interstitial markings, lung/pleural calcifications.

†Prominent pulmonary vasculature, atelectasis, high hiatus.

ND, not determined.

Table 3 Interobserver reliability among emergency department staff levels for any abnormality compared with senior radiologist interpretation

Emergency department physician training level	No (%) of chest radiographs read	κ	95% Confidence interval
Attending physician	113 (22.2)	0.31	0.20 to 0.42
Senior medical residents	103 (20.2)	0.38	0.28 to 0.48
Intermediate medical residents	136 (26.7)	0.35	0.26 to 0.44
Junior medical residents	90 (17.7)	0.47	0.36 to 0.59
Surgical residents	67 (13.2)	0.40	0.14 to 0.65

further evaluation the next day. The sensitivity for pleural effusion was very low (25.8%), yet only 6/31 (19.3%) effusions were of clinical significance, of which two were referred for further evaluation.

When categorised by levels of clinical significance (mild, moderate, and high as described in the methods section), the highest sensitivity of the emergency department physicians' interpretation was in the group with highly clinical significant radiographic findings (60%), with lower rates of sensitivity for the moderate significance group (31.7%), and low significance group (27.5%). There were no statistically significant differences between the observers at different levels of training ($p=0.87$).

Interobserver reliability, as assessed by the κ coefficient of agreement between all emergency department physicians and senior radiologists, was moderate to low: 0.40 (CI 0.35 to 0.46). There was no significant difference found for interobserver reliability among levels of the emergency department staff compared with the senior radiologist, as shown in table 3 ($p=0.33$).

DISCUSSION

Numerous studies have examined the interobserver reliability of radiographic interpretation in the emergency department. Each study was designed in a different manner thus making standardisation difficult.^{1 2 4 6 10 12-14} The trend over recent decades has been a decrease in the overall interpretation discrepancy rates to as low as 0.3%.^{3 4} Our study shows higher levels of misinterpretations as it was designed to maximise the potential discordance between emergency department physicians and radiologists by including all radiographic findings. Subtle signs, such as questionable consolidations, trace amounts of pleural effusion, very mild congestion, borderline cardiomegaly, etc, were regarded as positive findings therefore contributing to more errors of omission. Furthermore, basing the study on the actual medical records with no standard form indicates that many of these signs were not recorded in the discharge letter, either because of non-significance to the final discharge diagnosis, or because the findings were trivial to the specific patient (such as cardiomegaly in a known CHF patient). Using such a checklist or standard form might have increased the reported sensitivities, especially for findings determined to be of minor significance by the emergency department physicians. We also found that the percentage of errors increased as the clinical significance of errors decreased. Once again, this may be because emergency department physicians tend to be less meticulous in interpreting or recording findings which are likely to be of secondary or tertiary significance.⁸ Our study is limited by the fact that our "gold standard" did not represent a consensus of opinion, since 10 senior radiologists instead of one expert chest radiologist made the final interpretation of the radiographs. In our institution, as is common in many relatively small

community hospitals, the chest radiographs are read by various general radiologists, precluding comparison with other studies. We believe that if one trained chest radiologist had been reading the radiographs, the level of interobserver variability might have been higher.

We chose only chest radiographs of discharged patients and not all those done in the emergency department during the study period, whereas most studies examined all radiographs performed in the emergency department.^{1-4 6 10 12-14} Chest radiographs of admitted patients are more likely to have obvious positive findings, such as dense alveolar opacities, which are readily identified as pneumonia by even the most junior medical staff.⁷ The study was performed in the actual working environment of the emergency department and not in a "sterile", that is, purely theoretical or experimental setting,^{7 12 15 16} thus adding to the possibility of missing positive findings in the radiograph. As the emergency department physicians had to report their interpretation before the radiologist's consultation, we eliminated a bias found in some prior studies,^{3-5 10} which were based on only patients' charts for evaluation.

Chest radiography is a very commonly used investigation but detailed interpretation of the resultant film is relatively complicated.¹⁷ It has been reported that the chest radiograph is the most common radiograph to be misinterpreted by observers,^{15 16} especially in the setting of the emergency department.^{1-3 5 6} A study conducted to improve quality control showed that even when all radiograph misinterpretations were reduced, the percentage of chest radiographs did not differ.³ It is also noteworthy that our study consisted only of radiographs performed in adults, which usually have more abnormal findings than paediatric radiographs.^{6 9 10 12 13}

Radiograph evaluation entails subjectivity, variability, and uncertainty even when performed by experienced chest radiologists. Interobserver variability is difficult to overcome,^{8 9 15 16} and it should be emphasised that even between senior radiologists the κ coefficient has been shown to be low (0.45–0.65) for chest radiographs¹³ or specifically for pneumonia (0.22–0.52).¹⁵ Moreover, one study reported that 17% of 352 radiographs read by emergency department physicians and considered "misinterpreted" by radiologists, were later found accurate.¹⁸ Our study showed intermediate levels of emergency department physicians' reliability at all levels of training compared with senior radiologists. Interobserver reliability was not evident as a function of training level. This result is not far from that described in previous studies,³⁻⁸ showing that attending physicians are slightly more accurate than residents. It has also been noted that attending physicians tended to be more confident but no more accurate than residents,² but to date, there are insufficient data to determine which radiographs are at a higher risk of being misinterpreted.^{2 14}

In conclusion, the only way to reduce emergency department physician's errors of interpretation is by teaching and improving interpretation skills, maybe as part of their residency, or by quality control measures.^{3 4} Cooperation between emergency physicians and radiologists, as well as having an efficient callback system when abnormalities are found, are mandatory.

Authors' affiliations

M E Gatt, G Spectre, Department of Medicine, Hadassah University Hospital, Mount Scopus, Jerusalem, Israel
O Paltiel, Hematology Unit and the Department of Social Medicine, Hebrew University-Hadassah Medical School
N Hiller, Radiology Department
R Stalnikowicz, Emergency Department

REFERENCES

- 1 Brunswick JE, Ilkhanipour K, Seaberg D, et al. Radiographic interpretation in the emergency room. *Am J Emerg Med* 1996;14:346–8.

- 2 **Mayhew FE**, Rust DD, Aldag JC, *et al.* Accuracy of interpretation of emergency department radiographs: effect of confidence levels. *Ann Emerg Med* 1989;**18**:826–30.
- 3 **Preston CA**, Marr JJ, Amaraneni KK, *et al.* Reduction of “callbacks” to the ED due to discrepancies in plain radiograph interpretation. *Am J Emerg Med* 1998;**16**:160–2.
- 4 **Espinosa JA**, Nolan TW. Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. *BMJ* 2000;**320**:737–40.
- 5 **Kuritzky L**, Hardy RI, Curry RW. Interpretation of chest roentgenograms by primary care physicians. *South Med J* 1987;**80**:1347–51.
- 6 **Walsh-Kelly CM**, Melzer-Lange MD, Hennes HM, *et al.* Clinical impact of radiograph misinterpretation in pediatric ED and the effect of physician training level. *Am J Emerg Med* 1995;**13**:262–4.
- 7 **Young M**, Marrie TJ. Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. *Arch Intern Med* 1994;**154**:2729–32.
- 8 **Herman PG**, Hessel SJ. Accuracy and its relationship to experience in the interpretation of chest radiographs. *Invest Radiol* 1975;**10**:62–7.
- 9 **Herman PG**, Gerson DE, Hessel SJ, *et al.* Disagreement in chest roentgen interpretation. *Chest* 1975;**68**:278–82.
- 10 **Fleisher G**, Ludwig S, McSorley M. Interpretation of pediatric X-ray films by emergency department pediatricians. *Ann Emerg Med* 1983;**12**:153–8.
- 11 **Fleiss JL**. The measurement of interrater agreement. *Statistical methods for rates and proportions*. 2nd Ed. New York: John Wiley & Sons, 1981: 212–36.
- 12 **Kramer MS**, Roberts-Brauer R, Williams RL. Bias and “overcall” in interpreting chest radiographs in young febrile children. *Pediatrics* 1992;**90**:11–13.
- 13 **Simon HK**, Khan NS, Nordenberg DF, *et al.* Pediatrics emergency physician interpretation of plain radiographs: is routine review by a radiologist necessary and cost effective? *Ann Emerg Med* 1996;**27**:295–8.
- 14 **Lufkin KC**, Smith SW, Matticks CA, *et al.* Radiologist’s review of radiographs interpreted confidently by emergency physicians infrequently leads to changes in patient management. *Ann Emerg Med* 1998;**31**:202–7.
- 15 **Albaum MN**, Hill LC, Murphy M, *et al.* Interobserver reliability of the chest radiograph in community acquired pneumonia. *Chest* 1996;**110**:343–50.
- 16 **Robinson PJA**, Wilson D, Coral A, *et al.* Variation between experienced observers in the interpretation of accident and emergency radiographs. *Br J Radiol* 1999;**72**:323–30.
- 17 **Nyboe J**. Evaluation of efficiency in interpretation of chest X-ray films. *Bull World Health Organ* 1966;**35**:535–45.
- 18 **O’Leary MR**, Smith MS, O’Leary DS, *et al.* Application of clinical indicators in the emergency department. *JAMA* 1989;**262**:3444–7.

ECHO

Using rifampicin against TB risks a rare reaction

Please visit the Postgraduate Medical Journal website [www.postgradmedj.com] for link to this full article.

A case study of pneumonia induced by rifampicin in pulmonary tuberculosis should remind doctors to be alert to this very rare—but potentially fatal—reaction.

The case was of an 81 year old man being treated in hospital for pulmonary tuberculosis of the left lung with rifampicin (0.3 g/day), isoniazid (0.3 g/day), and ethambutol (0.5 g/day). Seven days after starting treatment he developed fever and dyspnoea, and x ray and computed tomography examinations showed shadows on the opposite, right, lung.

The man’s white blood cell count had increased from 6070/μl to 13 540/μl, comprising 81% neutrophils, 6% lymphocytes, 11% monocytes, and 3% eosinophils. Bronchoalveolar lavage fluid contained 83% lymphocytes, 14% neutrophils, and 4% macrophages; the ratio of CD4/CD8 lymphocytes was 10.5. A drug stimulation lymphocyte test (DLST) showed a stimulation index of 370% with rifampicin, 170% with isoniazid, and 130% for ethambutol.

Tuberculosis treatment was withdrawn. Steroids were given to correct progressive respiratory failure, and a gradual recovery followed. Streptomycin treatment (0.5 g/day) was started on day 26. Isoniazid and ethambutol treatment resumed three months after they were stopped, and the combined treatment continued for six months with no adverse reaction.

The only other published case has suggested that it is rifampicin toxicity that induces pneumonia, but the positive DLST result here suggests stimulation of the immune system may be another way.

Rifampicin is a valuable drug for treating tuberculosis. Its well known complications include hepatitis, fever, and blood problems; pneumonia is a rare, but dangerous, reaction.

▲ *Thorax* 2002;**57**:1000–1001.