

Assessing the accuracy and certainty in interpreting chest X-rays in the medical division

I Satia, S Bashagha, A Bibi, R Ahmed, S Mellor and F Zaman

ABSTRACT – The chest X-ray (CXR) is an important diagnostic tool in diagnosing and monitoring a spectrum of diseases. Despite our universal reliance on the CXR, our ability to confidently diagnose and accurately document our findings can be unreliable. We sought to assess the diagnostic accuracy and certainty of making a diagnosis based on 10 short clinical histories with one CXR each. We conclude from our study that specialist registrars (StRs) and consultants scored the highest marks with the highest average certainty levels. Junior trainees felt least certain about making their diagnosis and were less likely to be correct. We recommend that StRs and consultants review all the CXRs requested to ensure accuracy of diagnosis. There also needs to be discussion with the Joint Royal Colleges of Physicians Training Board (JRCPTB) about the need of including a separate CXR competency as part of a trainee’s generic curriculum on the e-portfolio, something which is currently lacking.

KEY WORDS: Chest X-ray, radiology, reporting, training, education

Introduction

The chest X-ray (CXR) is an important diagnostic tool in diagnosing and monitoring a spectrum of diseases - from managing the acutely unwell patient with pneumonia, heart failure or pneumothorax, to investigating suspected lung cancer, tuberculosis (TB) or interstitial lung diseases. It is the most widely performed X-ray and is considered to be the most difficult to interpret.¹ All grades within the general internal medicine (GIM) team are ever reliant on the CXR, yet our ability to confidently diagnose the plethora of pathophysiology or to accurately document our findings can be unreliable.

Comparison between physicians and radiology reporting has suggested radiologists provide improved quality² and accuracy³ of reporting. However the reporting of CXR for medical patients admitted to hospital vary across the UK. Due to resource limitations, our local hospital policy is that all admission X-rays are

reported, but this may not be available at the time of senior medical review on the medical assessment unit. This is not an uncommon scenario in other NHS trusts due to staffing levels as highlighted by the Royal College of Radiologists report, *Clinical radiology: a workforce in crisis*.⁴ Therefore, for all other CXRs the local policy states:

In accordance with the requirements of the The Ionising Radiation (Medical Exposure) Regulations (I(R)MER) under local agreement, the reporting responsibility for this examination has been delegated to the referring practitioner whose clinical report will be directly entered in to the patient notes. If a formal report or opinion is required then please contact the radiology department.

Furthermore there are no separate formal curriculum competencies for trainees, nor is CXR reporting systematically assessed in undergraduate or post-graduate examinations for the core or general internal medicine trainee. The current curriculum does, however, contain workplace based assessments (WBA) which include a comments box on ‘able to order, interpret and act on investigations’, but I would argue that this is a rather limited and non-specific method of assessing competency in interpreting CXRs. Likewise, medical finals and MRCP part 2 examinations may have some CXRs to interpret but it does not cover the breadth of diagnoses to gain competency and certainty in diagnosis. In the absence of capacity from radiologists, this area deserves further attention.

We therefore conducted an initial small retrospective audit to assess if disagreement in reporting between general physicians and radiologists would have an impact on the medical management of a patient. We retrospectively compared the reporting of 50 CXRs by the medical physician with that of a consultant radiologist. The overall concordance of reporting was 78%. There were 21/50 normal CXRs of which the physician correctly reported 20/21 (95%) as normal. The remaining were either acutely abnormal (14/50, 28%) or had chronic abnormal changes (15/50, 30%). Only 6/14 (43%) acutely abnormal CXRs were reported correctly and 11/15 (73%) reported on chronic changes. There were no CXRs which were reported as normal by the physicians when there was abnormal pathology. This suggested that although physicians correctly identified normal CXRs, they made the correct acute diagnosis only 50% of the time, and especially struggled to differentiate between pneumonia and heart failure.

The results of this small audit prompted a more thorough study of the diagnostic accuracy and certainty of interpreting CXRs of all grades of doctors within the medical division.

I Satia,^{1,2} clinical research fellow; S Bashagha,¹ foundation year 1; A Bibi,¹ foundation year 1; R Ahmed,¹ foundation year 1; S Mellor, consultant radiologist; F Zaman,¹ consultant chest physician

¹Royal Blackburn Hospital, East Lancashire Hospitals NHS Trust, UK;

²Education and Research Centre, University Hospital of South Manchester, Manchester, UK

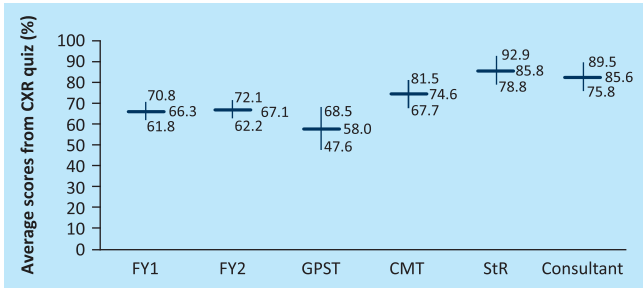


Fig 1. Comparison of average scores from CXR quiz across grades with 95% confidence intervals. CMT = core medical trainee; CXR = chest X-ray; FY = foundation year; GPST = general practitioner trainee; StR = specialist registrar.

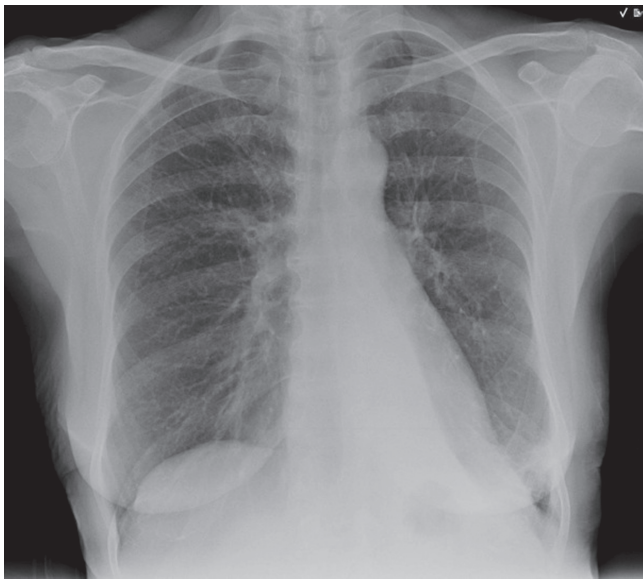


Fig 2. Left lower lobe collapse.

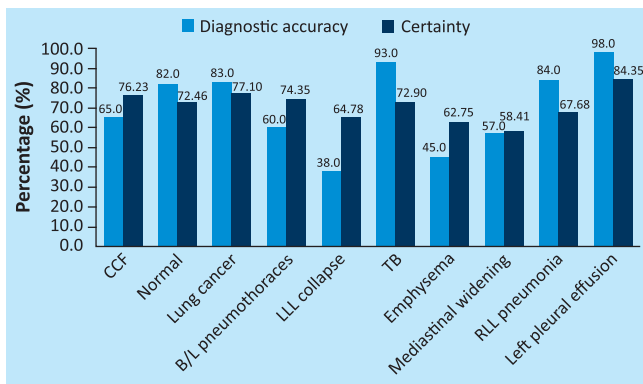


Fig 3. Comparison of diagnostic accuracy with certainty for each CXR diagnosis averaged across all grades. B/L = bilateral; CCF = congestive heart failure; CMT = core medical trainee; CXR = chest X-ray; FY = foundation year; GPST = general practitioner trainee; LLL = lower left lobe; RLL = right lower lobe; StR = specialist registrar; TB = tuberculosis.

Aim

To assess the accuracy and confidence of CXR reporting by all grades within the medical division for common chest diagnoses presenting to the medical assessment unit.

Method

10 CXRs were presented to all grades within the medical division in a large rural district general hospital with a short clinical history and one digital CXR which was agreed by a consultant chest radiologist and consultant respiratory physician. The CXR quiz was performed in groups of the different grades at different times. Doctors were asked to give a pre-test certainty (out of five) in reporting CXRs and then also give individual certainty levels for each of the CXR diagnoses. No multiple choices of differentials were given and each group of doctors were given 10 minutes to complete the quiz and were all marked by one respiratory registrar. A feedback of results and explanation of each CXR diagnosis was given in a medical audit meeting after all doctors had completed the quiz. There were no exclusion criteria.

The 10 CXR diagnosis that were shown were: congestive cardiac failure, normal, lung cancer with ribs metastases, bilateral pneumothoraces, left lower lobe collapse, left upper lobe TB, chronic emphysema, mediastinal widening, right lower lobe pneumonia, and left pleural effusion.

Results

A total of 138 doctors completed the CXR quiz: 46 foundation year 1s (FY1), 38 foundation year 2s (FY2), 10 general practitioner trainees (GPST), 13 core medical trainees (CMT), 12 specialist registrars (StR) and 19 consultants, of which four were respiratory physicians (Table 1). The average results across all grades with 95% confidence intervals are shown in (Fig 1). Within each grade data of doctors, data followed a normal distribution. The FY1s and FY2s had similar average scores of 66.3% and 67.1% respectively with similar confidence intervals suggesting that all foundation trainees were at a similar level of accuracy. Although there was a dip in the average score of the GPSTs to 58%, this could be accounted for by only 10 participants some of whom have come from surgical or psychiatry backgrounds. Thereafter, there seemed to be a trend towards higher scores in the CMTs (74.6%) and StRs (85.8%), while consultants scored equally well (82.6%).

On comparison of accuracy scores between foundation trainees (FY1/FY2) and StRs/consultants there is a statistical difference ($p=0.0002$) between those groups, but not with the GPSTs ($p=0.152$) or CMTs ($p=0.056$). In addition there is a significant difference between CMTs and StRs ($p=0.036$), but not with consultants ($p=0.113$). The GPSTs had the lowest score but with the widest confidence intervals, however, there was a statistical significance on comparison with CMTs ($p=0.019$), StRs ($p=0.0005$) and consultants ($p=0.0013$).

Analysing the range of scores show similar levels at foundation trainee level, which becomes narrower towards the StR group,

with a minimum score of seven and maximum of 10. There was an unexpected dip of a minimum score of 5 in the consultant group.

The least well-answered CXR diagnoses were left lower lobe collapse (38%) (Fig 2), emphysema (45%), and mediastinal widening (57%), whilst the best answered diagnosis were left pleural effusion (98%), TB (84%) and right lower lobe pneumonia (83%). However, there were some noteworthy points: 35% of doctors were unable to differentiate between heart failure and pneumonia, 18% were unable to diagnose a normal CXR, 17% were unable to spot the 3 cm right apical mass and 55% were unable to recognise features of chronic emphysema. For the latter, doctors were writing multiple diagnoses including emphysema, pneumonia and pulmonary fibrosis, and so were given zero marks. Interestingly, 18% of doctors were also able to notice an apical pneumothorax unilaterally but failed to spot the pneumothorax on the opposite side – termed ‘satisfaction of search’.

To assess how certain doctors were at making their diagnosis, each CXR was then compared with the corresponding average percentage certainty of diagnosis for all doctors (Fig 3). Doctors were on average over-certain in the incorrect CXRs they answered; left lower lobe collapse (38% score vs 64% certainty), emphysema (45% vs 62%), bilateral pneumothoraces (60% vs 74%). Interestingly, for the correct diagnosis, the average certainty for diagnosing effusions, TB and pneumonia were lower. This suggests the certainty of diagnosis varied; uncertainty with the correct diagnosis and over-certainty with an incorrect diagnosis.

Discussion

This is an important study of both the diagnostic accuracy and certainty of CXR interpretation, which to our knowledge, has not been performed here in the UK across all medical grades. We have demonstrated better diagnostic accuracy and certainty as the training grade increases. Mehrotra *et al*⁵ compared CXR competency of clinicians from medicine, accident and emergency (A&E), anaesthetics, intensive therapy unit (ITU), surgery and radiology. Consultants and registrars attained significantly higher scores than junior doctors, specialists (consultants and registrars in radiology and respiratory medicine, n=7) achieved significantly higher scores than non-specialists. A similar study in the US has shown similar results comparing medical students, interns, residents, and fellows;⁶ an increased level of training was associated with overall higher score. However, in both these studies accuracy of diagnosis was not assessed alongside certainty.

Jeffrey *et al*⁷ did assess certainty and accuracy of 52 final year medical students in the UK. The median score in their study was 12.5/20 and the overall degree of certainty was low. On no radiograph were more than 25% of students definite about their answer. It was suggested that students had received little formal radiology teaching (2–42 h, median 21) and had accounted for such low certainty and accuracy figures.

Table 1. Summary of results.

Grade	Average pre-test certainty (out of 5)	Average score (%)	Maximum score (%)	Minimum score (%)
FY1 (n=46)	2.9	66.3	90.0	30.0
FY2 (n=38)	3.2	67.1	90.0	30.0
GPST (n=10)	3.1	58.0	90.0	40.0
CMT (n=13)	3.5	74.6	100.0	50.0
StR (n=12)	3.5	85.8	100.0	70.0
Consultant (n=19)	4.1	82.6	100.0	50.0

CMT = core medical training; GPST = general practitioner trainees; FY = foundation year; StR = specialist registrar

We therefore contacted two local medical schools regarding the delivery of radiology teaching for medical students, and were told that radiology does form a very important part of the curriculum but the responsibility of delivering this was for the host hospital to arrange. One medical school did however provide opportunity for students to learn via e-learning modules. Students would then be shown some radiological images in either the OSCE or in the written exam.

Like many trusts, our own local trust is responsible for providing radiology teaching to its own foundation trainees, which it had done with one formal general radiology session for two hours in one year. The local deanery said it provided between two and seven hours of a broad general radiology teaching for its CMTs in two years but there was no radiology teaching for StRs in GIM. Although surprising, this may reflect teaching sessions trying to cover the trainee’s curriculum. CXR interpretation is not a separate skill as part of the foundation, core medical or a StR trainee’s e-portfolio. This is a training issue which needs discussion if we are going to produce the next generation of general internal medicine consultants able to provide a quality service. Otherwise, we may be more reliant on radiologists in reporting all our CXRs, the capacity for which, may not be available.

Despite the importance of this small study, it does have its limitations. Firstly, each diagnosis could have a spectrum of disease manifestation on CXR and simply being able to diagnose right lower lobe pneumonia on one CXR is not an indication of competency for all pneumonias. The same is the case for lobar collapses or lung nodules. Secondly, when comparing average certainty of each CXR diagnosis, all the scores from all grades were merged to create one value and then compared with the corresponding sum of all average diagnostic accuracy scores across all grades. Given that there were more foundation trainees, this average value may be negatively skewed and more representative of junior trainees rather than senior trainees. It should also be noted here that the certainty score (out of five) is a subjective qualitative assessment for each doctor compared against an objective quantitative score from a correct diagnosis or not. Hence, it may not be completely accurate or reliable when trying to compare the two different

forms of results on one graph (Fig 3). Certainty scores also had a much narrower range of scores and this probably reflects the subjective nature of self-assessment. It may be worth trying a more reliable and accurate method of assessing certainty using descriptive terminology rather than numerical values and possibly with lesser scales, for example, 'uncertain', 'certain', and 'definitely certain'.

Another criticism could be that we have assumed the CXR diagnosis from a chest radiologist and chest physician to be in our control with 100% sensitivity and specificity for interpreting those 10 CXRs. I would argue that for the 10 CXRs shown, the specificity would be as high as 100% given the classical nature of the diagnostic image. However, had CT scans of the thorax been performed of those 10 patient at the time of the CXR, other abnormalities not seen on the CXR could have been missed. The other possibility is that other radiologists and chest physicians ought to have formed a bigger control group.

Finally, the results in Fig 1 also need some careful consideration. As there is an upward trend towards better diagnostic accuracy and certainty towards the higher grades, one might conclude that the current system of training must be working and it might be that the additional years of experience contributed towards better scores. However, on comparing StR and consultant scores, there was no significant difference in scores. In fact, if we ignored the scores of the four respiratory physicians in that consultant group, the average score would have dropped significantly from 82.6% to 73%. It would be unfair to draw too many conclusions from a study of only 19 consultants and 12 StRs, but clearly more vigorous and highly-powered detailed studies and analysis need to be undertaken looking at the training and reporting mechanisms at other hospitals across the UK.

Conclusion

Specialty registrars and consultants scored the highest marks with the highest average certainty levels. Junior trainees felt least certain about making their diagnosis and were less likely to be correct. We recommend that StRs and consultants must review all the CXRs requested to ensure accuracy of diagnosis. There also needs to be discussion with the JRCPTB about including CXR competency as part of a trainee's generic curriculum on the e-portfolio, something which is currently lacking.

References

- 1 Kanne JP, Thoongsuwan N, Stern EJ. Common errors and pitfalls in interpretation of the adult chest radiograph. *Clin Pulm Med* 2005;12:97–114.
- 2 Weiner SN. Radiology by non-radiologists: is report documentation adequate? *Am J Manag Care* 2005;11:781–5.
- 3 Al aseri Z. Accuracy of chest radiograph interpretation by emergency physicians. *Emerg Radiol* 2009;16:111–4.
- 4 Royal College of Radiologists. *Clinical radiology: a workforce in crisis*. London: RCR, 2002.
- 5 Mehrotra P, Bosemani V, Cox J. Do radiologists still need to report chest X-rays? *Postgrad Med J* 2009;85:339–41.
- 6 Eisen LA, Berger JS, Hegde A, Schneider RF. Competency in chest radiography. A comparison of medical students, residents, and fellows. *J Gen Intern Med* 2006;21:460–5.
- 7 Mehrotra P, Bosemani V, Cox J. Do radiologists still need to report chest X-rays? *Postgrad Med J* 2009;85:339–41.
- 8 Jeffrey DR, Goddard PR, Callaway MP, Greenwood R. Chest radiograph interpretation by medical students. *Clin Radiol* 2003;58:478–81.

Address for correspondence: Dr I Satia, Education and Research Centre (Level 2), University Hospital of South Manchester, Southmoor Road, Manchester, M23 9LT. Email: imransatia@gmail.com