

Project Title: Performance Testing for Radiologists Interpreting Chest Radiographs

Purpose::

The aim was to develop a system to assess the image interpretation performance of radiologists in identifying signs of malignancy on chest radiographs.

Methods

A test set of 30 digital chest radiographs was chosen by an experienced radiologist consisting of 11 normal and 19 challenging abnormal cases. The abnormal cases all had biopsy proven pathology; the normal cases had at least two years of imaging follow up. 14 radiologists with a range of experiences were recruited. Participants individually read the test set displayed on a standard reporting workstation, with their findings entered directly onto a laptop running specially designed software. For each case they were given the relevant clinical information and were asked to mark any perceived abnormality and rate their level of suspicion on a five point scale (normal, benign, indeterminate, suspicious or malignant). On completion of the test, participants were given instant feedback had the opportunity to review cases where there was disagreement with the expert opinion and pathology. The time taken for the participants to complete the test was recorded.

Differences between the participants' performance were assessed using ROC analysis..

Results

The experience of the participants in reporting chest radiographs ranged from 1 to 26 years (Mean = 9 years, Mdn = 5 years). Participants' performance (ROC score) varied significantly between two groups (6 post-fellowship consultants, and 8 radiology residents). Radiology residents' performance as measured by ROC score was significantly poorer compared to post-fellowship consultant radiologists ($M_{RS} = 0.76$, $M_{PFC} = 0.93$, $p = .003$). There was a positive correlation between image interpretation performance (ROC Mean = 0.85, SD = 0.11) and years of reading experience (Mean = 9, SD = 8.58), $r = .573$, $p < .05$, $n = 14$.

There was a trend for radiology residents to take longer to complete the task (Mean = 26.51s) compared to post-fellowship consultant radiologists (Mean = 19.65s), but this did not quite reach statistical significance ($p = .07$).

Conclusion

This pilot study demonstrates that it is possible to devise a method for performance testing the reporting of chest radiographs.

Clinical Relevance Statement:

Chest radiographs are the first line imaging test for patients with chest symptoms suspicious of malignancy, this pilot study demonstrates that it is possible to devise methods to test performance of the reporting radiologist.