Methods employed for chest radiograph interpretation education for radiographers: A systematic review of the literature

Keywords: radiography, chest, education, training, image interpretation, reporting

Introduction

Role progression in image interpretation within radiography can reduce patient waiting times, ensure patient safety and decrease costs.\(^1\,\text{,}\,2\) Chest image interpretation is a challenging and skilful task; the large variation in patient anatomy, the range of pathologies which can present on a chest image and the appearance of different pathologies add to the complications arising when undertaking this task.\(^3\) Initial training and education can provide a solid base to address these complications and familiarise interpreting clinicians with them.\(^4\) The role of radiographers within chest image interpretation varies greatly with Preliminary Clinical Evaluation (PCE) and clinical reporting roles now available in the United Kingdom (UK).\(^4,\,5\) On reviewing the limited evidence available on performance by reporting radiographers, chest image interpretation mean specificity and sensitivity was reported to be 95.4\% and 95.9\%, respectively.\(^4\) This result was similar to that of the top 20 radiologists reported by Potchen et al. (2000) where the area under the curve of the Receiver Operating Characteristic (ROC) analysis was estimated to be 0.95.\(^6\) However, the evidence on the education of radiographers and competence within this relatively new field of chest image interpretation by radiographers is limited. Little is known about alternative training methods available, merits of how this training is undertaken and outcome measures of tested training methods. A review of current education/training provided for radiographers on chest image interpretation is required to assess whether radiographers are being adequately trained and whether the methodology employed can influence the accuracy of radiographers in this area of practice.

Methods

A systematic literature review was performed by searching the following healthcare databases: Medline (1949-present), Pubmed (1947-present), Cumulative Index to Nursing and Allied Health (CINAHL) (1937-Present), the Cochrane Library Database (1974-Present) Scopus (1823-Present) and Embase (1980-Present). The “Medical Subject Heading” (MeSH) was used to identify related keywords. The search strategy was developed using variations of the following keywords: radiographer, radiologic technologist/technician, x-ray, image, film, radiograph, chest, thorax and axial.

Articles were included if they were in English, focused on chest radiograph image interpretation, involved radiographers as participants and featured a form of training in the interpretation of chest radiographic images. Articles were excluded if they featured a modality other than plain chest imaging, or if they were articles on the imaging examination, dose, quality or technology, were case specific or focused on patient safety and care/service evaluation.

The lead author reviewed all abstracts and identified papers which met the inclusion criteria. The other authors independently screened these papers to ensure they met the inclusion criteria. All authors met to compare findings; any differences in reviewers' judgments were resolved through discussions until consensus was reached.
Data was extracted by the lead author using a predesigned form and this data was entered into the results tables.

For the purpose of the review, the quality of the studies were assessed based on a variation of the questions provided in the Critical Appraisal Skills Programme Oxford UK (CASP) tools for a cohort study and diagnostic study. The combination was used as no suitable alternative was available for these mixed methods papers. If the answer to a question was 'yes' the article was scored 1, if the answer to a question was 'can't tell' or 'no' a score of 0 was awarded for that question.

Results

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart summarises the literature review search results (Figure 1). A total 645 articles were identified within the searches; following application of the inclusion/exclusion criteria, 13 articles were reviewed. A summary of the articles is provided in Table 1. The quality of the studies was high ranging from scores of 6-10. Studies scored lower if they gave little information on the recruitment process, only one person acted as the reference standard, if they failed to list the confounding factors of their studies or focused solely on the detection of one chest pathology.

A total population of 649 participants were assessed on their chest image interpretation accuracy between the years 1978-2016. Of these participants, 466 were students or experts within the radiography or reporting radiography profession or equivalent. Just over 30% of the studies were completed within the last six years (2010-2016), featuring approximately 69% (445/649) of overall participants and approximately 69% of radiographer (students and qualified personnel) participants (323/466). The sample size within studies varied greatly. The smallest study featured one reporting radiographer. The largest studies featured 148 and 134 radiography students or experienced radiographers. The numbers of participants within the other studies ranged from 1 to 40. A total of 10 studies were conducted in the United Kingdom (UK), where the role progression of chest image reporting by radiographers is established, one study was conducted in Africa, one within America and one within South Pacific countries (to include participants from Fiji, Soloman Islands, Vanuatu, Cook Islands, Kiribati and Niue).

The measurements recorded by each study are provided in Table 1. One study utilised a questionnaire and recorded participants' confidence levels; an increase in confidence was seen post training in all groups, with the mean confidence increasing from 7.71/10 to 8.73/10 (median confidence value of seven pre tutorial to nine post tutorial).

As demonstrated in Table 1, there was a large variability of participant experience and age. The level of expertise varied greatly; 144 students, 279 radiographers and 43 reporting radiographers comprised the 466 radiography participants featured within the studies.

The accuracy measurements presented in the literature are summarised in Table 2. There was a large variance in how the accuracy/performance measurement was presented within the literature: figure of Merit (FOM) was presented by two studies; ROC (n=1); Alternative Free-Response Receiver Operator Characteristic (AFROC) (n=4); average accuracy/average
agreement score \( (n=5) \). Mean diagnostic accuracy was recorded as the lowest pre-training by Semakula-Katende\(^{11}\) as 47.3% and as the highest pre-training by Flehinger\(^{16}\) as 87.7%. The lowest post-training accuracy was noted by Cowan\(^{17}\) as 5 participants scored under 50% and the highest post-training accuracy percentage was 92.5%.\(^{11}\) The agreement between the reference standard (image interpretations by expert(s) taken as the standard) and radiographers ranged from 86.7% to 96%.\(^{4,9}\) Specificity and sensitivity \( (n=5) \) were recorded. Sensitivity values ranged from 53.5% \(^{11}\) to 87.8% (A/E images) \(^{14}\) pre-training and from 79.1% \(^{16}\) to 100% (A/E images) \(^{14}\) post-training. Specificity values ranged from 40% (A/E images) \(^{14}\) to 72.1% \(^{11}\) pre-training and from 59.9% \(^{16}\) to 98.8% \(^{18}\) post-training. Studies asking participants to allocate images to a designated group had the lowest sensitivity and specificity values pre-training.\(^{14,13}\) Positive predictive power ranged from 67.7% \(^{14}\) to 78.4% post-training \(^{18}\) and negative predictive power ranged from 96.3-100% post-training \(^{14}\) \((n=2)\) False positives (FP): False Negatives (FN), True Positives (TP) and True Negatives (TN) or a combination of the four outcome measures were recorded \((n=3)\). Percentage of FN's decreased from 47% to 42% following training.\(^{19}\)

There was a wide variation between the training interventions applied across the studies (Table 2). Overall there was a positive effect on participant performance recognised within each study radiographers' accuracy increased or was at a high level following training. However, one study noticed no effect and a slight decrease in performance when participants were provided with time to preview the image before their interpretation or when asked to ‘free search’ the image.\(^{10}\) Six articles featured radiographers who attended postgraduate training in chest image interpretation. Two articles supplied feedback either through eye tracking, an image preview session or an overlay of eye tracking onto an image as their intervention.\(^{10,16}\) Five studies provided lectures, presentations or an apprentice programme to participants, of which three incorporated feedback/instruction sessions into the training.

The eye tracking metrics were measured in six of the studies under review (Table 3); however specific data on this outcome measure was only supplied within two articles.\(^{10,15}\) Donovan\(^{15}\) assessed eye tracking metrics of participant groups pre- and post-feedback/no feedback, whereas Litchfield presented figures on the eye tracking metrics within the feedback provided to participants. Litchfield\(^{15}\) used various eye tracking movements such as those from novices/experts, and presented the eye tracking metrics relative to these rather than using the eye tracking to assess participants' performance as in Donovan.\(^{15}\) In general, time to first fixate (measure of how long it took before a test participant fixated on an area of interest (AOI)), fixation count (measure of the number of times the participant fixated on an AOI), average fixation length and gaze time decreased when participants were given a second look at the image (post-feedback/no feedback) compared to their first look at the image.\(^{15}\) There was no significant effect between pre and post "no feedback" condition and a significant difference between the pre and post "feedback" condition. By studying the eye tracking metrics, it was concluded that expert and naïve participants were less affected by feedback.\(^{15}\) Therefore, perceptual feedback may be beneficial for those in early training.

Six studies restricted the time allocated to participants. Image viewing time was limited to 40 minutes \(^{20,21}\) or one hour \(^{10,19,22}\) for all images, or 30 seconds per image within the training.\(^{11}\) Authors briefly mentioned the use of a time limit to reduce the effect of fatigue on participant performance.\(^{21,2}\) One study gave an estimate of the study taking approximately one hour to
complete (Litchfield et al. 2010). Two studies provided figures for the mean decision/scrutiny times per cohort. Training decreased the time taken to diagnose from a mean 33.9s to 31.1s (p=0.02). However, feedback provided increased and decreased decision times depending on the type of feedback and participant group within Litchfield et al. (2010).

The instructions and guidance given to participants varied greatly across the articles and impacted the answer and quality of the answer. Two radiographers were asked to screen the images within a department and to focus on the presence/absence of lung cancer. In another study, participants were asked to use a search strategy in their interpretation and given a choice of four pathologies as their diagnosis. Elsewhere, participants were tasked with highlighting the pathology with the well-known ‘red dot’ approach. Articles (n=6) named the pathology and asked participants to determine if images were normal/abnormal and give details on position of the pathology if they identified the image as abnormal. Other authors (n=3) provided participants with choices of response such as three/four choices with which to assign to each image.

There was a difference in which form the chest image was presented to the participant (i.e. film based radiograph or digital format image) (Table 4). Three studies identified their use of a particular age group within the test bank. Four articles acknowledged their use of digital/digitised chest images. Few articles tested the participants on their ability to identify more than one pathology type (n=4). The remaining articles (n=7) focused either on the presence of lung cancer or pulmonary nodules.

Authors presented the ratio of normal or abnormal images used within their image test banks (Table 4). Three articles focused on images which were encountered in clinical practice and the prevalence rate of abnormal images was calculated on completion of the study. The participants were allocated a specific task when viewing the images in clinical practice within an allocated time period. Prevalence rates are an important consideration when testing participants on medical image interpretation. This is seen within Pusic et al’s study where high sensitivity was seen within a group tested using a high number of abnormal images (0.69±0.24) compared to groups trained with medium (0.63±0.21) and low (0.51±0.24) number of abnormal images. This was converse to the high specificity seen within the group trained using a low number of abnormal images (0.83±0.10) compared to groups tested with medium (0.70±0.15) and high (0.66±0.17) numbers of abnormal images.

Many studies featured within the review failed to give information on how many experts formed their reference standard (n=6). Articles featured a radiologist/consultant radiologist (n=4), three consultant radiologists (n=2) or six consultant radiologists (n=1) as their reference standard (Table 4).

Discussion
A broad and informative systematic review has been performed identifying 13 high quality articles detailing the provision of chest image interpretation education to radiography students, radiographers or reporting radiographers and assessing their performance before and after the intervention. The evidence indicates that, although a wide range of training types were employed, radiographers demonstrated high and improved levels of accuracy where chest
image interpretation training was undertaken. The large majority of the studies conducted within the UK (approximately 77%) is unsurprising given the role progression of chest image interpretation by reporting radiographers was established in England, UK.\textsuperscript{5,12,13} These findings supply supporting evidence for the role extension of radiographers into chest image interpretation providing the appropriate support, assessment and audits are carried out to ensure these accuracy levels are upheld.\textsuperscript{4}

The role progression of chest image interpretation by reporting radiographers in recent years correlates with the number of articles recently produced on radiographers’ performance in chest image reporting.\textsuperscript{4,9} Articles published outside the last decade have focused on participants allocating images to a given category or highlighting abnormal images.\textsuperscript{16,18} Nonetheless, the most recent article focused on radiography in South Africa provided participants with choices of response with which to assign to each image;\textsuperscript{11} this fits with the clinical expectations of radiographers within that country. Those articles featuring chest reporting are based within the UK and arose following the establishment of reporting radiographers within this practice area. Beyond the difference in interpretation task, approximately 77% (10/13) articles were from the UK.

Imaging equipment has changed over recent years due to healthcare demands, technological advances and safety regulations. ‘Film’ is suggestive of the older imaging equipment techniques used but may have been a simple lapse in terminology used by Donovan\textsuperscript{15} or indeed may indicate that they opted to use the older imaging display format. Flehinger\textsuperscript{16} was the only study in the review to include lateral chest radiographs and the only study undertaken in America; the additional lateral chest radiographs are probably reflective of the imaging protocols of the country or the year which the study was carried out (1978).

The information given (i.e. film presented over a light box or a digital image displayed on a high quality viewing monitor) could have impacted the participants’ performance and ability to distinguish between absence/presence of a pathology. Viewing conditions and image presentation could impact the ability of the viewing medium to deliver sufficient spatial and contrast resolution of the image.\textsuperscript{27,28} Digitised images, as used within five of the studies,\textsuperscript{17, 19, 20, 21, 22} could have impacted the interpretation process and the participants’ ability to identify the pathology. Ability change and visual system function explored by Garland\textsuperscript{29} and Kundel\textsuperscript{30} showed little difference in observer performance and both agreed that imaging diagnosis is fallible despite the many years and digital advances between these studies.

This review has demonstrated that accuracy is higher in the studies where specific postgraduate training is being evaluated.\textsuperscript{4,9} This is because a programme has been designed specifically to cover all requirements of the reporting radiographer’s chest image interpretation role and will directly influence patient care. Other roles undertaken by radiographers in chest image interpretation such as red dotting or training attempt to improve the radiographer’s general ability to complete Preliminary Clinical Evaluation (PCE) and to identify pathologies but do not carry this higher level of responsibility and liability.\textsuperscript{12,13} Nonetheless, only two studies provided evidence within this field, one of which contained one participant. As such, there remains to be a lack of knowledge on the current training and standards for PCE.

Few studies tested the participants on their ability to identify and distinguish between a range of pathologies (n=3). Only two recent studies tested participants on this task both of which
were investigating the effect of postgraduate education. There have been no recent training studies, other than those featuring postgraduate education, which tested participants on their identification of a range of chest pathologies. Monitoring interpretation time and setting fixed interpretation time has its benefits in avoiding fatigue; however, setting a time limit may impact upon the participants. Although studies, which include postgraduate programmes, provide evidence of high levels of accuracy, there has been no detailed investigation into training that may complement these programmes. The eye tracking feedback approach was used in lung nodule detection only and with little or no guidance provided with eye movements. In Hughes paper at least 25/26 radiographers acknowledged tutorials incorporating a search strategy as useful; however, many (n=19) claimed to have a technique for looking at chest radiographs already and, therefore, the value of the search strategy implemented may be misunderstood. Tutorials, lectures and short courses proved to be useful training methods; however, these are applicable for chest image interpretation roles rather than reporting and have yet to be tested for their effect within chest image reporting by radiographers. Postgraduate training demonstrated high levels of accuracy from all participant groups, to a standard of performance similar to radiologists. Training/feedback with little or no instruction is not beneficial to trainees, and the lack of guidance supplied with eye tracking feedback may have led to it being little or no use as the participants doubted their ability if shown the image for a long pre interpretation time period. Education aids may be crucial in supporting development of reporting pathways and skills where postgraduate training may not be an option; this may be particularly useful for radiographers within developing countries. The results from this review signify a high performance by radiographers who receive training to complete chest image interpretation, this evidence provides support to undertake these roles elsewhere and to maximise the use of resources, particularly in developing countries. Implementation of electronic checklists may be an option to reinforce or support training. The majority of such checklists proved successful during a review by Kramer, with only 1/15 articles identified as non-beneficial. As far as we are aware, such electronic checklists have not been tested specifically for their use in chest image interpretation by radiographers. Guidance and instructions given to participants when undertaking the study varied greatly. This ultimately affected the task given to participants within the study. Where participants are given a specific pathology to identify or a list of groups to assign an image to, it is questionable as to whether the participant has readily identified the pathology or categorisation options influenced their choice. Tasks where these options have been provided may be easier than forming a diagnosis on the image alone as the provision of specific groups/pathologies may make the participant cautious and mindful of these during their image viewing. The number of experts acting as the reference standard varied greatly within studies. Verification bias, work up bias and incorporation bias have been identified as considerations in relation to application of the reference standard. Verification bias relating to whether all of the images interpreted by the observers were also interpreted/accessed by the reference standard; work-up bias being a result of the observer’s report being used to decide whether the reference standard is applied; and, incorporation bias being whether the observer’s report was used to generate the reference standard. These biases must be considered when
scrutinising the reference standard within each study. A majority of the reviewed studies tested participants on images which were each interpreted by a reference standard and which was one or more consultant radiologist, therefore, reducing the interference of these biases. However, six studies failed to provide information on the reference standard details. One radiologist is a failing in good reference standard generation. Approximately 31% of the studies within the review had more than one radiologist/reading stated as the reference standard.

Limitations
Accuracy measures were presented in many formats and training varied greatly within the studies. Therefore, limited comparisons can be made and limited conclusions can be taken. The non-specific search terms used may have been a possible limitation of the review. There is a possibility that different more precise terms might have provided more relevant papers; however, the search terms were chosen to lessen the risk of missing important articles. Only English written articles were included within the review, potentially allowing key articles relevant to the review to be excluded. We requested additional information from lead authors; however, some studies continued to have little information on data collected.

Conclusion
Radiographers demonstrate high and improved levels of accuracy where chest image interpretation training has been undertaken. Strong evidence from high quality studies indicates that accuracy improved regardless of the training type; however, some methods appeared to enhance this improvement more than others. Training techniques should be combined to maximise learning and chosen to suit the skill of chest image interpretation the radiographer wants to learn. It is advisable that whichever training method is chosen that it is accompanied by monitoring of student performance to ensure its worthiness, validity and success.
References:


5. Canterbury Christ Church University. PgC clinical reporting (adult chest) validation document; 2002.


Figure 1: Summary of literature review search records using PRISMA group flow chart (2009)

Records identified through database searching
\( n = 642 \)

Additional records identified through other sources
\( n = 3 \)

Records after duplicates removed
\( n = 358 \)

Records screened
\( n = 358 \)

Records excluded by title or abstract
\( n = 299 \)

Articles matching exclusion criteria
\( n = 45 \)

Full-text articles assessed for eligibility
\( n = 58 \)

Article included as experiment within another article
\( n = 1 \)

Studies included in analysis
\( n = 13 \)