Introduction

The Preliminary Clinical Evaluation (PCE) is a commenting scheme designed to improve the specificity of the widely adopted red-dot abnormality detection system; the Society and College of Radiographers[1] are advocates of this system and the Standards for Proficiency outline that radiographers should be able to distinguish abnormal appearances and trauma processes (HCPC 2013). Furthermore, there is an expectation that all radiographers have sufficient knowledge of radiographic anatomy and common abnormalities (Education and Career Framework for the Radiography Workforce document (SOR 2013), which would facilitate effective participation in a PCE system. PCE provides radiographers with an opportunity to have a positive impact on timely patient management. Effective communication of abnormal findings is considered to reduce the time-to-diagnosis, which may also have an impact on the length of hospital stay[2]. Despite recognised benefits, there has been minimal publication of large-scale empirical studies confirming the success of PCE. The uptake of PCE has been slow with the suggestion that this may in part be due to the increase of reporting radiographer activity[3]. If PCE is to be a worthy successor to the red-dot abnormality detection system, radiographers must provide a service that is accurate, and an effective driver of improved patient outcomes.

The meta-analysis by Brealey et al[4] suggests radiographers have good accuracy when using a red-dot abnormality detection system, albeit against varying reference standards with associated differential verification biases. Very little exists by way of objective observer studies that assess performance but a few recent studies aptly illustrate the image interpretation abilities of radiographers.

Piper and Paterson[5] undertook an alternative free-response receiver operating characteristic (AFROC) study to assess the effect of training on the ability of 38 participants (radiographers and nurses) to accurately locate an abnormality and to simply state the nature of the abnormality. Improvements were observed after training with radiographers demonstrating post-training increases in figure of merit (0.63 to 0.73), sensitivity (60% to 69%), and specificity (73% to 83%), respectively.

The FROC study by McEntee and Dunnion[6] indicated that radiographers can accurately detect abnormal wrist images with sensitivity comparable to that of radiologists (radiographers 87.7%, radiologists 88.9%), but specificity is poor (radiographers 64.4%, radiologists 80.5%). McEntee and Dunnion[6] concluded that, although not statistically significant, the number of years of experience could positively affect interpretation skill; they did not however assess the effects of training on performance. Earlier work by Hardy & Culpan[7] has proven that sensitivity and specificity levels do improve following training; 72% to 88% and 50% to 53%%, respectively.

It is generally accepted that an increasing number of years of radiographic experience will have a positive impact on the correct interpretation of trauma images. In less experienced staff it is likely that providing training for newly qualified radiographers would expedite accurate contributions in a PCE system.

Despite claims of good accuracy, it is thought that PCE has not been widely implemented due to a perceived lack of confidence and inadequate training[2,8] with previous research suggesting that the requirement to provide a written comment caused a reduction in abnormality detection accuracy[7, 9]. However, this is not a universal opinion, where it has been suggested that good red-dot performance indicates an ability to provide a written comment[10]. If training issues do exist, and are not addressed appropriately, then the effectiveness of the PCE could be restricted[7].

Much of the previous work discussing the uptake of PCE focuses on the quality of training and the preparedness of radiographers to provide an accurate PCE comment. Graduate radiographers are expected to have sufficient image interpretation ability, despite a lack of certification of competency[9]. The aim of this paper is to evaluate the fracture detection performance and PCE accuracy of a small sample of graduate radiographers using an objective observer study to assess detection accuracy, and a scoring system to assess commenting accuracy. Given that questions remain about training and the ability of radiographers to provide a comment, this study will operate a pre- and post-training design to assess the impact of focussed training on a graduate radiographer's ability to accurately localise and describe a red-dot type abnormality.

Materials & Methods

Local Research and Development, and the Health Research Authority[11] decided that the project was suitable as service evaluation. The clinical cases selected were all acquired more than 12-months prior to this study. This reduces the likelihood of new fractures being detected on our review of the cases, since the patient is likely to have presented symptomatically in this time period if an occult fracture had been present. This was important to ensure the correct fracture status in normal and abnormal images. Where follow-up imaging was available, it was reviewed to ensure that no occult fractures were present on cases used in the observer study. All observers provided written consent.

Case Selection

A three-month audit of abnormality prevalence for all examinations of trauma to single appendicular parts was undertaken in the study centre revealing a 29.4% incidence of abnormality. We used this data to determine the number of normal/abnormal cases (prevalence) for the observer study, and also the distribution of appendicular examinations that should be included. The range of the subtlety of abnormalities within the selected cases was also consistent with the local workload. One of the authors (BS) compiled the caseload based on the findings of the abnormality prevalence audit. Replicating the local clinical workload provides a comparative assessment of participant interpretation, relative to their clinical practice(12). We performed a sample size calculation to predict the required number of cases, based on six observers completing the study. Obuchowski(13) developed a mathematical model to provide sample size tables for ROC analyses based on the intricate relationships of accuracy, inter-observer variability, patient variability and the correlations in accuracy imposed by the study design. Test alpha was set at 0.05 to control the probability of Type I error, while the power is set at 80%. We estimated that 58 cases would be required for a suitably powered study with a ratio of 4:1 (negative: positive) cases. This ratio was the nearest to the 29.4% prevalence of abnormal cases established from our audit. The image bank of 58 examinations consisted of 17 abnormal appendicular examinations and 41 normal appendicular examinations. Cases containing normal variants were not excluded and were considered as normal. The mean distribution of each appendicular examination over the previous three months was calculated alongside the percentage occurrence. The percentage occurrence was then applied to the sample size to provide the number of each examinations required. Table 1 summarises the 17 abnormal cases and the gold standard PCE comments, and the 41 normal cases used in this study. The gold standard PCE descriptions are a consensus of two Advanced Practitioner's interpretations; who verified the descriptions of the abnormalities rather than relying on the report. DICOM headers were removed from all cases to ensure anonymity. All annotations identifying fractures or dislocations were also removed. Each abnormal case contained only one abnormality to allow quantification of a single comment. No discrepancies with the original radiological report were identified in the case selection process.

Case	Fracture Location (Score 3: Side, Bone, Location)	Fracture Type (Score 1)	Movement (Score 1)
1	Left Radial Head	Intra-articular	Minimal Displacement
2	Left Scapula (Lateral)	Comminuted	Posterolateral Displacement
3	Right Distal Radius	Buckle	Dorsal Angulation
4	Left Distal Tibial Epiphysis (Lateral)	Longitudinal	Anterior Displacement
5	Left 2nd Proximal Phalanx (Base)	Oblique	Minimal Displacement
6	Left Distal Radial Metaphysis	Buckle	Dorsal Angulation
7	Right Glenohumeral Joint	Dislocation	Posterior Displacement
8	Left Proximal Tibial Metaphysis	Incomplete	Undisplaced
9	Left 5th Metatarsal Base	Transverse	Undisplaced
10	Right 3rd Metatarsal Neck	Stress	Undisplaced
11	Left Distal Radial Metaphysis	Buckle	Dorsal Angulation
12	Left Proximal Metaphysis Proximal Phalanx	Longitudinal	Undisplaced
13	Right Lateral Malleolus	Oblique	Minimal Displacement
14	Right 5th Metacarpal Base	Oblique	Undisplaced
15	Left 4th Proximal Phalanx Neck	Oblique	Lateral Displacement
16	Right 1st Toe Interphalangeal Joint	Dislocation	Plantar Displacement
17	Right 5th Metacarpal Neck	Oblique	Volar Angulation
18 to 58	Normal Cases: Ankle (x7) Elbow (x3) Femur (x1) Finger (x3) Foot (x4) Forearm (x1) Hand (x4) Humerus (x1) Knee (x4) Scaphoid (x1) Shoulder (x5) Tibia (x1) Toe (x1) Wrist (x5)	N/A	N/A

Table 1: Breakdown of the image case mix used showing the gold standard PCE comment for each of the abnormal images.

Observer Performance Study & PCE Scoring

Four observers evaluated the 58 cases on two occasions: (i) pre-training and (ii) post-training. All observers were in a preceptorship period; eight weeks of training elapsed between the two evaluations. We based our sample size calculation on 6 observers, but only 4 were able to complete the study. For one of the observers it transpired that they did not fulfil the inclusion criteria (newly-qualified radiographer, first-appointment), and for another there was an unavoidable delay in commencing their employment, therefore they were excluded from the study. An eight-week training schedule, separating the pre- and post-training evaluations, consisted of intensive educational sessions designed to deliver information relative to abnormality detection. The sessions were designed and delivered by one of the authors (BS), Advanced Practitioner (skeletal reporting). The introductory session covered basic terminology and concepts, which familiarised participants to a systematic approach of detecting a fracture, forces and fracture patterns, established vocabulary, and a model of forming a comment. All appendicular body parts were covered; each session followed the same format, which included radiographic anatomical knowledge, common fractures, assessment lines and measurements, concepts relative to each body part and the relevant abnormal cases, as well as examples to practice forming a comment.

All observers were trained to use the software for the observer study and how to approach the study. They were given a test set of 10 images with which they were asked to localise suspicious areas and provide a PCE comment. This test-set could be repeated until the observer was confident with the data collection method. Each case could include 2-4 images, depending on the type of examination. Observers were instructed to mark all areas suspicious of fracture/dislocation with a mouse click; this prompted an unmarked slider-bar rating scale to appear with which they could indicate confidence (1-10) in their decision. Moving the slider further to the right indicated increased confidence. Since multiple images were available for localisation (i.e. AP and lateral), it was possible that a fracture could be localised on more than one image. In such cases, we took the highest rating, as only one rating could be used per fracture/dislocation in the analysis. It was not necessary for the observers to mark the fracture on all projections for it to be deemed a successful localisation. An acceptance radius classified observer marks; and a visual assessment confirmed whether mark-rating pairs were true or false. All image evaluations were completed on a 20" LCD flat panel monitor at 60Hz (NEC MultiSync LCD 2090UXI, 600 x 1200, NEC Display Solutions, Itasca, Illinois, USA) using ROCView[14] to record observer responses. Each image evaluation was completed in a different randomised order.

For each localisation the observers were also asked to provide a PCE comment. Pre-training comments were based on experience from undergraduate education. Post-training they were expected to be familiar with the components of an accurate PCE comment, following the eight week training programme. They were scored on the following components, with each assigned a single point for a maximum score of 5 for each comment: name of bone, location of fracture, anatomical side (L/R), fracture type, and the presence of any movement, such as displacement or angulation. A gold standard comment was agreed by two experienced musculoskeletal reporting advanced practitioners.

Statistical Analysis

We are interested in the accuracy of the clinical comment and the precise localisation of abnormalities. The equally weighted jack-knife alternative FROC JAFROC (wJAFROC) figure of merit is sensitive to location information and defines probability that a true abnormality is rated with higher confidence than a false localisation[15]. Data was analysed using Rjafroc; an implementation of wJAFROC analysis in the R programming language. A difference in abnormality detection between pre- and post-training was considered significant if the result of the overall F-test was significant and the 95% confidence interval (CI) did not include zero. Test alpha was set at 0.05.

Results

A significant difference in fracture detection performance was found between pre- and post-training evaluations for a fixed reader random case analysis (F (1,57) = 10.57, p = 0.0019). The reader averaged wJAFROC FOM and 95% CIs for pre- and post-training were 0.619 (0.516, 0.737) and 0.703 (0.622, 0.852) respectively. The reader averaged wJAFROC curves are displayed in Figure 1. All readers demonstrated improvement from pre- to post-training, as evidenced by the increase in wJAFROC FOM, Table 2.

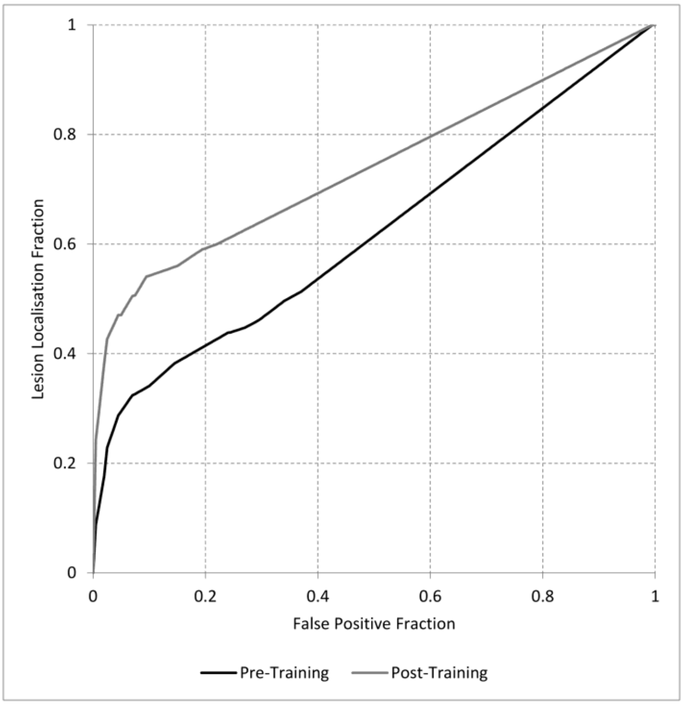

Figure 1: The observer averaged wAFROC curves for pre- and post-training image evaluations.

Reader	Pre-Training wJAFROC FOM	Pre-Training PCE Score	Post-Training wJAFROC FOM	Post-Training PCE Score
1	0.680	13	0.789	39
2	0.570	18	0.730	31
3	0.662	29	0.684	28
4	0.564	8	0.742	26
Mean	0.619	17	0.737	31

Table 2: Comparison of each reader's pre- and post-training wJAFROC FOM and PCE scores.

Abnormality (fracture or dislocation) detection was assessed on a case-by-case basis for the 4 readers in this study to identify further training needs. Reader averaged detection rates improved from pre- to post-training, 42% and 56% respectively. From these cases, it was apparent that these novice observers had difficulty in detecting cases with undisplaced fractures (cases 8, 10, & 12). None of the readers could detect these abnormalities post-training. Another trend was observed for distal radius fractures in paediatric patients, where each fracture (cases 3, 6, & 11) was only successfully localised by one reader. There was a 50% reduction in false localisations after training.

The PCE score was composed of five criteria; bone, location, side (L/R), fracture type, and movement. Table 3 illustrates the increases in each of the PCE criteria following the training period. A paired t-test was used to compare the pre- and post-training PCE scores. This demonstrated a statistically significant improvement in PCE comment for all observers, $t(4) = 9.68$, $p = 0.0006$, mean (95% confidence interval) 11.20 (7.99,14.41). In cases where the fracture was not localised the PCE score was generally consistent with this event; however, it was still possible to achieve a PCE score if the precise site had been missed (i.e. indicating the correct anatomical side). Additionally, in some cases in the pre-training evaluation the PCE score was still low even when the fracture had been successfully localised.

Scoring Criteria	**Total PCE Score (All Observers)**		**Score change between pre and post test**
	Pre-training	Post-training	
1 – Correct Bone	23	34	+ 11
2 – Correct Location	19	34	+ 15
3 – Correct Side (L/R)	15	23	+ 8
4 – Fracture Type	6	18	+ 12
5 – Displacement/Angulation	5	15	+ 10
Total	68	124	+ 56

Table 3: The total PCE score of all observers in pre- and post-training evaluations. The table indicates the total score for each of the five criteria, pre- and post-training score, and the change between pre- and post-training score.

Discussion

We found a statistically significant improvement in fracture detection as a result of a focused 8-week training programme. We have also been able to demonstrate an improvement in precision when using a PCE comment as a result of this training. If a PCE commenting system is to be successfully introduced then the radiographers using this system must demonstrate equal, if not better performance when compared to that of the previously used red dot system. There is great potential for success of a PCE system, as it can reduce the ambiguity that can be caused by a non-location sensitive 'red-dot' system.

The increases in performance we observed following the training phase of the study substantiates the study by Hardy & Culpan[7] that assessed 115 radiographers' abilities to recognize and describe radiographic abnormalities following attendance at a red dot study day course. Their results showed that following training, red dot sensitivity and specificity improved alongside abnormality description. Further correlation is seen with the findings of Piper and Paterson[5] who also reported increases in performance following training; despite their significant findings it was concluded that further work is needed to evaluate performance in image interpretation.

Detection rates increased for all but one reader. Interestingly, this reader (3) produced a very similar PCE score in both pre- and post-training. This may indicate a difference in undergraduate education, as their pre-training score was much higher than the other readers. However, the 50% reduction in false localisations reveals that the intensive training sufficiently improved the reader's ability to recognise normal appearances, echoing the work of Wright & Reeves[16]. The overall improvement in PCE score from pre- to post-training was evident in all of the 5 criteria used to score the comment; with the greatest improvement (score +15) observed in the description of the correct type of fracture. This improved appreciation of fracture morphology is recognised as providing benefits in diagnosing and managing the patient[17].

Two participants correctly localised and described a fracture of the second proximal phalanx on the PA wrist projection (case 5) in the post-training test compared to zero participants in the initial test. This suggests improvement in the overall search of the image. Discussion of the satisfaction of search phenomenon should be included in any training program;

whereby the detection of one abnormality interferes with detection of another, and is often affected by knowledge of common fractures[18]. This level of understanding may not manifest itself in the search strategy of newly qualified radiographers.

In this study we have a trend of a failure to detect buckle fractures of the paediatric distal radius, and this correlates with the findings of previous work[19]. There were also difficulties in detecting subtle and undisplaced fractures; all of these findings could help direct training for newly qualified radiographers. We recommend that intensive PCE training should be included in the preceptorship program or during the transitional period from graduate to independent practitioner. It must be stressed though that the issue of sustaining any improvements in performance is just as challenging as attaining the desired level. Previous work by Mackay (2006) indicated that the immediate improvements in abnormality detection following training were not demonstrable after 6 months; reinforcing the need for regular CPD sessions to maintain standards, not just for newly qualified radiographers but also those who are more experienced. For the newly qualified radiographer the transition from student to practitioner can be quite daunting. However, the pressure of contributing successfully to a PCE system can be reduced by this comparatively simple, cheap and regular departmental training intervention.

This study has demonstrated the effectiveness of the method we proposed; the study should now be repeated with a larger sample size and over a larger number of cases in order to generalise the results to the population of newly qualified radiographers. However, the initial results are encouraging, where we have demonstrated the effectiveness of a focussed training programme to improve fracture detection rates and the accuracy of a PCE comment. Experiential learning, peer support and educational reading cannot be excluded as potential influences on the performance increase from pre- to post-training evaluations, but it would not be practical to conduct this study in isolation of any these external factors. As with all observer studies using a test/re-test method there is a risk of memory effects influencing the second evaluation. However, the 8-week period between evaluations, randomisation of image order and the fact that the observers would see a large number of other clinical cases during this time as part of their daily work do limit this effect. Another limitation of this work is the relatively small sample of observers and the fact that the clinical cases, and estimation of fracture prevalence, were drawn from a single centre. However, we believe the methods applied to be robust, but would be strengthened by a multi-centre approach. The sample of observers was reduced from our original calculation; this will have a negative impact on the power of the study.

Future work could also assess the impact of the accuracy of a PCE comment on emergency practitioners' evaluation of the image, and the speed and appropriateness of care delivered to the patient as they return to the emergency department.

Conclusion

This study found a statistically significant improvement from pre- to post-training fracture detection performance. Post-training PCE scores also showed an overall increase. These results were also consolidated by a 50% reduction in false localisations post-training. A larger, multi-centre study, using a greater number of observers should be conducted to provide a result that can be generalised to the population of UK radiographers. However, on the basis of these findings we recommend an intensive training program would benefit newly qualified radiographers in providing the necessary framework for participating in a PCE system.

Conflict of Interest

No conflicts of interest influenced this work.

References

1. SOR. Preliminary Clinical Evaluation and Clinical Reporting by Radiographers : Policy and Practice Guidance. 2013.
2. Lancaster A, Hardy M. An investigation into the opportunities and barriers to participation in a radiographer comment scheme, in a multi-centre NHS trust. Radiography [Internet]. Elsevier Ltd; 2012;18(2):105–8. Available from: http://dx.doi.org/10.1016/j.radi.2011.08.003
3. Snaith B, Hardy M. Radiographer abnormality detection schemes in the trauma environment-An assessment of current practice. Radiography. 2008;14(4):277–81.
4. Brealey S, Scally A, Hahn S, Thomas N, Godfrey C, Crane S. Accuracy of radiographers red dot or triage of accident and emergency radiographs in clinical practice: a systematic review. Clin Radiol. 2006;61(7):604–15.
5. Piper KJ, Paterson A. Initial image interpretation of appendicular skeletal radiographs: A comparison between nurses and radiographers. Radiography [Internet]. Elsevier Ltd; 2009;15(1):40–8. Available from: http://dx.doi.org/10.1016/j.radi.2007.10.006
6. McEntee MF, Dunnion S. A FROC analysis of radiographers performance in identification of distal radial fractures. Eur J Radiogr Eur J Radiogr. 2009;1(3):90–4.
7. Hardy M, Culpan G. Accident and emergency radiography: A comparison of radiographer commenting and ‘red dotting’. Radiography. 2007;13(1):65–71.
8. Neep MJ, Steffens T, Owen R, McPhail SM. A survey of radiographers’ confidence and self-perceived accuracy in frontline image interpretation and their continuing educational preferences. J Med Radiat Sci [Internet]. 2014;61(2):69–77. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4175834&tool=pmcentrez&rendertype=abstract
9. Hardy M, Snaith B. Radiographer interpretation of trauma radiographs: Issues for radiography education providers. Radiography [Internet]. Elsevier Ltd; 2009;15(2):101–5. Available from: http://dx.doi.org/10.1016/j.radi.2007.10.004
10. Coleman L, Piper K. Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers. Radiography [Internet]. Elsevier Ltd; 2009;15(3):196–202. Available from: http://dx.doi.org/10.1016/j.radi.2007.12.001
11. Health Research Authority. Is my study research? [Internet]. 2016 [cited 2016 Jul 3]. Available from: http://www.hra-decisiontools.org.uk/research/
12. Hardy M, Flintham K, Snaith B, Lewis EF. The impact of image test bank construction

on radiographic interpretation outcomes: A comparison study. Radiography [Internet]. Elsevier Ltd; 2016;22(2):166–70. Available from: http://dx.doi.org/10.1016/j.radi.2015.10.010

13. Obuchowski N. Sample size tables for receiver operating characteristic studies. AJR Am J Roentgenol. 2000;175(3):603–8.
14. Thompson, J. D., Thompson, S., Hogg, P., Manning, D., and Szcepura K. ROCView : prototype software for data collection in jackknife alternative free-response receiver operating characteristic analysis. 2012;85(September):1320–6.
15. Chakraborty D, Berbaum K. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys. 2004;31(8):2313–30.
16. Wright C, Reeves P. Radiography Image interpretation performance : A longitudinal study from novice to professional. Radiography [Internet]. Elsevier Ltd; 2016;6–12. Available from: http://dx.doi.org/10.1016/j.radi.2016.08.006
17. Chew FS. Skeletal Radiology : the Bare Bones. Philadelphia: Wolters Kluwer Health; 2012.
18. Ashman C., Yu J., Wolfman D. Satisfaction of Search in Radiology. 2000;(August):541–4.
19. Nunn H, Nunn DL. Determination of difficult concepts in the interpretation of musculoskeletal radiographs using a web-based learning/teaching tool. Radiography [Internet]. Elsevier Ltd; 2011;17(4):311–8. Available from: http://dx.doi.org/10.1016/j.radi.2011.06.006