Canterbury Christ Church University's repository of research outputs

http://create.canterbury.ac.uk

Please cite this publication as follows:


Link to official URL (if available):

https://doi.org/10.1016/j.radi.2019.04.013

This version is made available in accordance with publishers' policies. All material made available by CRenaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: create.library@canterbury.ac.uk
Multi-professional image interpretation: performance in preliminary clinical evaluation of appendicular radiographs

Lockwood P, Pittock L.

Clinical and Medical Sciences Research Hub, School of Allied & Public Health Professions,
Canterbury Christ Church University, Kent, UK

Introduction

Within the United Kingdom (UK) various approaches to radiographic abnormality detection systems (RADS) have been in clinical practice since the start of the 20th century. The current preliminary clinical evaluation (PCE) guidance formulated by the Society and College of Radiographers (SCoR) is defined as a commentary on imaging findings to referrers. The application of PCE is not intended as a substitute for a definitive report but as a form of RADS. The objective of PCE is to support the referrer’s clinical judgements to aid initial patient treatment and management for acute injuries.

The Health and Care Professions Council (HCPC) Standards of proficiency: Radiographers requires radiographers to “be able to distinguish between normal and abnormal appearances evident on images (13.14), understand the structure and function of the human body in health, disease and trauma, as well as common pathologies and mechanisms of disease and trauma (13.21), be able to formulate specific and appropriate management plans (14.2) be able to undertake and record a thorough, sensitive and detailed clinical assessment” (14.6). The SCoR promotes the view that PCE is a core part of radiographers’ scope of practice, and should be underpinned by education and competency development. The application of PCE into the scope of practice of other healthcare professions in the acute setting requires workforce development and professional body guidance documents specific to individual roles.

At present, studies have shown the potential for radiographers to produce PCEs, with limited up to date literature demonstrating the image interpretation ability of other healthcare professionals involved with the treatment and management of patients requiring radiographic imaging for trauma. The College of Paramedics definition of the role of an advanced paramedic practitioner, in line with the Royal College of Emergency Medicine ACP pathway for managing patients through accident and emergency (A&E) and urgent care pathways, includes requesting radiographic examinations and image interpretation. Similar to the role of emergency nurse practitioner (ENP) role that has been long established in the NHS. Likewise, the multidisciplinary response to emergency department demands has also included the development of the role of emergency physiotherapists to triage, manage and discharge patients, extending the scope of practice to include requesting and interpreting radiographic imaging. Drivers for PCE implementation include Department of Health (DoH) priorities to improve healthcare efficiency in the National Healthcare Service (NHS) against workforce shortages and raising patient demand. In assessing the NHS, the Care Quality Commission and the DoH have advised that to continue delivering sustainable healthcare delivery the NHS must support local improvements in quality of care and patient outcomes.

Image interpretation of radiographs for healthcare professionals is now beginning to be embedded by UK universities as part of the multidisciplinary clinical assessment skills of Advanced Clinical Practitioners (ACP) working in the acute setting. The short course at (Anonymised for peer review
University is a part-time interprofessional healthcare non-credit bearing short course for practitioners who interpret radiographic findings, influencing immediate patient management. The learning and teaching provided the skills necessary to produce PCE commentaries for trauma imaging of the appendicular skeleton. The campus-based study included image interpretation workshops (12 sessions) covering radiographic anatomy, pathology, osteology, normal variants, and search strategies with lectures on descriptive terminology, common errors, auditing, legal and professional issues by academic staff, reporting radiographers and consultant radiologist and acute medicine physician clinical practice supervisors. Supplementary sessions were provided for participants without a radiography background to introduce principles of radiographic techniques to recognise poor technical adequacy (positioning and exposure) which may impede interpretation. Educational support was provided by web-based radiographic image banks, on-line anatomy and pathology resources for off-campus viewing and clinical practice learning agreements to gain experience and exposure to prospective workloads. The assessment involved formative PCE commentaries completed in the participant’s clinical workplace and a summative Objective Structured Examination (OSE) on campus.

Aims

This paper evaluates the educational programme OSE results of three cohorts of multidisciplinary healthcare professionals’ interpretation of appendicular radiographic examinations at the end of the PCE module. The null hypothesis proposes no difference exists in the means of the test scores in the set of given observations, regardless of the professional backgrounds (variables of previous experience and training) as all the participants attained the same learning and teaching. The opposite may be shown through the observer performance that there is a difference in the means of the test scores (alternative hypothesis).

Methods

Approval for this study was received from the institutional ethics and governance approval panel. A sample size calculation was not performed for recruitment of participants in the study design (to gain a power (1-β) and effect size) as this study reports the results of students enrolled on a programme of study (fixed sample size).

Literature review of established standards

A literature search was completed to establish standards for comparison of the results of each subgroup (profession) in image interpretation of appendicular radiographic examinations (index test) of traumatic injuries (target conditions). The selection criteria of search terms included the observer group ‘radiographer, nurse, paramedic, physiotherapist’ and the index test used ‘image interpretation, red dot, preliminary clinical evaluation’ with Boolean operator search terms. Databases applied included Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, CINAHL, PubMed Central and Google Scholar, filtering results to the period of 1985 to 2018. Additional studies were sourced through reference lists in found papers. Found studies were reviewed for methodological quality against the QUADAS-2 criteria. In this study the published data needed to be identifiable for inclusion (results as true positive (TP), true negative (TN), false positive (FP), false negative (FN) or sensitivity, specificity, agreement) or excluded if data were incomplete or missing. Synthesis of available study data was performed using MetaDiSc software and displayed in forest plots displaying pooled estimates for sensitivity, and specificity analysis. Additional statistical meta-analysis calculated Chi-square ($\chi^2$) with p-values ($p$), and inconsistency I-
square ($I^2$) to disclose any heterogeneity or consistency in the results, with degrees of freedom (df). Pooled log diagnostic odds ratio (DOR) was applied for any bias of study size and direction of effects. Final output measure calculated Summary Receiver Operating Characteristic (SROC) plots for each eligible subgroup result.

**Radiographic test bank**

The OSE caseload consisted of 25 retrospectively collected and anonymised appendicular skeletal radiographs (index test), with individual referral details of each case (9 females, age range 7-60 years; and 16 males, age range 4-57 years), presenting clinical information and referral source (12 from A&E, and 13 from general practice (GP). The abnormality prevalence consisted of 48% normal (including normal variant) cases and 52% abnormal cases (proportionate selection). Internal validity of the cases ensured a range of subtle and textbook traumatic injuries (target conditions) were applied (Table 1) to reflect clinical practice.

**OSE Reference standard**

The reference standard applied independent reporting (and blinded to the original radiologist report) of each OSE case by two reporting radiographers. The reports were reviewed for concordance of definitive answers (internal validity) and range of target conditions to reflect the level and ability of skills of the module learning outcomes by the programme team.

**Data collection**

The OSE was undertaken in an academic environment under controlled conditions using low-level lighting. Each radiographic case was displayed in Digital Imaging and Communications in Medicine (DICOM) format on KPACS software to allow full image adjustment (windowing, zoom, pan, scroll, etc. of the data sets) by the participants. The cases were displayed on DICOM compliant monitors (EIZO RadiForce RX350, 3MP, 54.1cm, 1536 x 2048 native resolution, 1,000 cd/m2 luminance, 1500:1 contrast ratio, DICOM Part 14 calibrated, with an anti-glare and anti-reflection coating) that complied with RCR image interpretation standards.

The OSE instructions briefed the candidates to provide a confidence level score for each case as normal or abnormal and provide a detailed free text description of findings. In abnormal cases, the exact location (side and anatomical site), primary condition (single or multiple) and any secondary complications (swelling, angulation, dislocation, etc.) were required.

**Test bank scores**

The OSE scoring system applied a maximum two marks for abnormal (13 cases) and one mark for normal (12 cases) if correctly classified (including location and description), a total of 38 marks from the 25 cases was available. Fractionated scores were used for cases of more than one abnormality present.

**Statistical analysis**

The programme team scored the candidate answer booklets to the reference standard answers. Correct answers were graded as TP or TN; incorrect responses were classified as FP or FN. Fractions (whole and partial) were applied in cases of multiple target conditions as defined in a previous
study\textsuperscript{46} to allow calculation for sensitivity and specificity.\textsuperscript{40} For individual reader performance Cohen’s Kappa (reader performance reliability) was applied for inter-reader agreement with 95% confidence intervals (95% CI).\textsuperscript{47} Subgroup (profession) analysis of OSE score, sensitivity, specificity with Analysis of Variance (ANOVA) for variation among and between the group’s performance was applied (alpha = 0.05). Alternative Free-response Receiver Operating Characteristic (AFROC) methodology\textsuperscript{48} was applied allowing for multiple pathological conditions and locations per case and Area Under the Curve (AUC) values estimated.

Results

Observer groups

The participants included nineteen nurses (N1-19; from ENP, GP and orthopaedic roles), and fourteen allied health professionals, seven radiographers (R1-7), five paramedics (P1-5) and two physiotherapists (PT1-2); Table 2. The demographics included twenty three females (age range 28-60 years, mean 40.4 years) and ten males (ages 28-48 years, mean 41 years). The clinical departments of the participants included four acute NHS district general hospitals and four minor injury NHS health centres with x-ray facilities.

Individual reader performance

The participant results demonstrated a range of individual test scores, sensitivity, specificity and Kappa outcomes (Table 2a-d). Participant group sizes varied which affected the range of scores for predication of overall mean analysis.

Subgroup test scores

The Kappa coefficient statistical measure of inter-rater agreement for the professional groups displayed a strong level of agreement for the physiotherapists (k=82), a good measure for the radiographers (k=72) and paramedics (k=61), and a moderate level of agreement for the nurses (k=56, Table 2a-d). The AFROC subgroup performance (Figure 1) displayed a limited range of confidence index ratings resulting in a constrained curve plotting of operating interval points. The subgroup AUC values were encouraging, with physiotherapists AUC 91.8 (95%CI 83-100), radiographers AUC 86.5 (95%CI 80.6-92.3); paramedics AUC 81.5 (95%CI 73.6-89.4), and nurses AUC 78.3 (95%CI 74-82.6).

The ANOVA estimated the variation (hypothesis testing) among and between the professional subgroups scores to calculate the differences among the sample means (table 3a-c). In this sample the exploratory data analysis result cannot reject the null hypothesis as ANOVA probability (p-value) for the OSE test scores (p=0.23) and OSE sensitivity score (p=0.65) were higher than the significance threshold (alpha=0.05), the OSE specificity score were equal but not lower (p=0.05) with the result not statically significant. To confirm this outcome the observed ANOVA values of F (Fisher–Snedecor distribution) against the F critical values (F-Crit), determined the F were less than the F-Crit (2.9) for the OSE test score, sensitivity, and specificity in all subgroup samples. A post-hoc power analyses on the sample size (to gain a power and effect size) would in effect repeat the p-value (ANOVA F-test confirmed result), in this situation the width and range of the 95% CI more appropriately reflects the statistical power (Tables 2a-d and Figure 1).
Subgroup performance comparison to literature

Comparison to published literature indicated a paucity of evidence from the subgroups which resulted in a wider literature search to include additional databases to confirm the limited pool of available studies for the nursing, paramedic and physiotherapist professions and provide a comprehensive review for scrutiny. A summary of the process and resulting studies are displayed in Figures 2, 3 and 4.

Synthesis of data included large-scale studies that involved multiple rounds of assessment (pre, mid, post, and follow-up data) which provided a wealth of quantitative results and relevant statistical information of performance ability. As such some studies were broken down into multiple entries (pre, mid, post, follow-up data) for the pooled assessment per professional group. The highest amount of eligible studies for inclusion on reader performance in musculoskeletal radiographic image interpretation was found in the radiography profession (41 studies from 1985–2018). Comparison of the pooled literature sensitivity 84% (95% CI 83-84%; Figure 3A) to this study 79.7% (95% CI 71.8-87.5%; Table 2a), pooled literature specificity 84% (95% CI 84-84%; Figure 3B) compared to this study 92.9% (95% CI 90.6-95.2%; Table 2a) and literature SROC AUC of 93.1 (Figure 3C) compared to the AFROC AUC 86.5 (Figure 1) in this study were encouraging. The pooled literature DOR (45.6; 95% CI 32.5-64.0; $x^2$ 3158.9; df 60, p=0.00; $I^2$ 98.1%) for the found studies displayed no concerning variation of DOR to represent publication bias.

The nursing profession evidence resulted in a small collection of eligible studies (from 1996-2014) which allowed extrapolation of results for analysis and construction of 2x2 contingency tables. Review of the pooled literature sensitivity 75% (95% CI 73-77%; Figure 4A) to this study 76.2% (95% CI 68-84.5%; Table 2b), and pooled literature specificity 80% (95% CI 77-82% Figure 4B) to the 80.4% (95% CI 75.1-85.8%; Table 2b) outcome in our study displayed strong results in this cohort performance. The literature SROC AUC of 91.2 (Figure 4C) in contrast to our study AFROC AUC of 78.3 (Figure 1) displayed minor variance. The results displayed in this study were equivalent to the found literature pooled results, the summary DOR (27.9; 95% CI 6.8-114.8; $x^2$ 348.8; df 8, p=0.00; $I^2$ 97.7%) for the literature did not suggest publication bias to be present.

Literature searching for skeletal radiographic image interpretation studies providing evidence for paramedic and physiotherapists abilities, identified no papers, thus limiting the comparison of these results to any standards for those professions.

Discussion

The participant results in this study were encouraging when compared to published studies assessing equivalent radiographic examinations and applying similar scoring systems. Radiographer image interpretation has been demonstrated extensively since the early work of Berman et al and the systematic review by Brealey et al, through to the work of Piper and Paterson and Coleman and Piper addressing a multidisciplinary approach and in recent studies of PCE ability of radiographers in the UK, South Africa and Australia.

The literature review found a disposition towards qualitative research from the nursing profession discussing the role of requesting and triage of radiographic imaging in the acute setting from the UK and Australia. The nursing studies identified echoed the concerns raised by Snaith and...
Hardy of an absence of details to assess study validity, highlighting a lack of radiology reporting (radiologist or reporting radiographer) participation and instead applying inadequate reference standards (image interpretations) of junior doctors, details of assessment methods, characteristics of cases, and omitted full disclosure of results and data.\textsuperscript{13,15,91,93,96}

Likewise, physiotherapists working in extended scopes of practice within the A&E setting to triage musculoskeletal injuries and request and interpret radiographs is a developing international practice (United States of America\textsuperscript{97}, Australia\textsuperscript{98-101}, UK\textsuperscript{102,103}, Ireland\textsuperscript{104}, Canada\textsuperscript{105}). But a similar theme was noted throughout these found studies that there was a lack of radiology reporting (radiologist or reporting radiographer) involvement in either training or assessing physiotherapists in radiographic image interpretation and diagnosis (or applied as an appropriate reference standard/ground truth in observer performance studies). Kersten et al\textsuperscript{106} explored the range of extended scope of practice of physiotherapists in the acute setting and raised similar concerns that due to the lack of professional guidance on radiographic image interpretation. Kersten et al\textsuperscript{106} advise the physiotherapy professional body may wish to consider prerequisites of formal training before participating in this extended role to ensure the quality of care for patients, regulation and protection for the practitioners. These concerns apply to all professions, and any healthcare professional working at an advanced level practice should be supported by professional frameworks and training for safe service provision. However, Kersten et al\textsuperscript{106} stopped short of explicitly highlighting radiology reporting (radiologist or reporting radiographer) input, or used as reference standards in training. More importantly, these studies fail to address the lack of rigorously applied methods to assess the ability of image interpretation\textsuperscript{40} which should be adequately evaluated before implementation of the role extension in clinical practice. These problems were also reflected in the paramedic\textsuperscript{107,108} literature.

Many variables potentially influence the results of the healthcare professionals enrolled in this study. The pre-existing skills and experience in physical clinical assessment of patients by nurses, paramedics and physiotherapists and utilisation of this knowledge can complement and enhance their clinical decision making in radiographic interpretation. Potentially this may provide an advantage over radiographers, although the undergraduate training, knowledge and experience at looking at radiographs, and the daily exposure and experience of this role in clinical practice by radiographers may likewise advantage them in the image interpretation task.

In examining the performance of the participants, errors occurred in the identification of upper limb cases containing multiple conditions. An example being an adult wrist with a fractured scaphoid (all noted) with an additional perilunate dislocation (missed by 16/33 participants, case 2). Likewise a paediatric elbow with raised fat pad and fracture of the medial epicondyle of the humerus (all noted) but with an additional fracture of the radial head (missed by 11/33, case 11). Adult shoulders also appeared in the FP errors with two normal cases having FP errors (6/33 case 4, 7/13 case 7), and the posterior gleno-humeral joint dislocation was incorrectly called anterior by 9/33 (case 9). Of the lower limb cases, the paediatric ankle case raised the most FN errors for missed avulsion fractures (8/33; case 8) and FP errors for a normal paediatric ankle (9/33; case 15).

Training of the autonomous emergency practitioner and ACP roles are establishing a generic collection of learning skills in clinical and diagnostic assessment, decision making in management and treatment plans, pharmacology and discharge. The inclusion of radiographic image
interpretation and PCE commentary training into ACP courses and CPD training for emergency healthcare professionals should attempt to train and assess individuals to an established benchmark score for safe clinical practice. Further research is recommended to investigate the utility of multidisciplinary PCE and image interpretation commenting downstream on decision making, time to treat, cost-effectiveness and patient outcomes.

**Limitations**

The cohort sample was small, and the particular professions were represented by inconsistent sample sizes (restricted paired data sets, producing inherent external validity bias) thus limiting the results outside of this study, a powered sample size is recommended for future studies. The literature data was not originally published in ROC methodology; thus we are aware caution should be applied comparing SROC AUC to AFROC AUC due to differences in calculations.

**Conclusion**

The comparison of image interpretation of appendicular radiographs in an academic environment by the radiographers and nurses to the published literature were encouraging (SROC AUC 93.1 and 91.2 respectively). The paramedics and physiotherapists results could not be compared to published standards or abilities on the task due to a lack of literature available. These small sample results, therefore, can provide an estimated baseline for future research. Radiographic image interpretation is now beginning to be embedded in multi-disciplinary ACP roles as an expectation by professional bodies, although with an absence of established standards for some healthcare professions to achieve in this task it raises potential regulatory, professional, and organisational safe practice concerns.

**Conflict of interest statement**

No conflict of interest, financial or otherwise, to declare.

**Acknowledgements**

We would like to thank all the healthcare professionals who participated in this study. The work described was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

**References**


36. The Care Quality Commission. Radiology review: A national review of radiology reporting within the NHS in England. The Care Quality Commission, Newcastle upon Tyne 2018


45. Royal College of Radiologists. Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices. 2012. London, UK.

46. (Anonymised for peer review manuscript).


50. Bowman S. Introducing an abnormality detection system by radiographers into an accident and emergency department: an investigation into radiographers’ concerns about the introduction of such a system. Res Radiography. 1991;1:2-0.
55. Ross C. The effectiveness of a short course of study on the ability of radiographers to highlight and describe abnormalities on skeletal accident and emergency films. MSc Thesis. South Bank University; 2002.


82. Rodriguez CM. A student radiographer’s accuracy in ‘red-dotting’ appendicular skeletal radiographs: Clinical Audit. 2017

83. Tay YX, Wright C. Development of Image Interpretation skills - a follow up study of Singapore diagnostic radiographers. 2017.


97. de Gruchy A, Granger C, Gorelik A. Physical therapists as primary practitioners in the emergency department: six-month prospective practice analysis. Physical therapy. 2015 Sep 1;95(9):1207-16.


Table 1. Test bank cases.

1. Normal Paediatric Tibia and Fibula
2. Abnormal Adult Hand [Fracture and Dislocation]
3. Abnormal Adult Hand [Fracture]
4. Normal Adult Shoulder
5. Abnormal Adult Knee [Fracture]
6. Abnormal Paediatric Tibia and Fibula [Fracture]
7. Normal Adult Shoulder
8. Abnormal Adult Ankle [Avulsion Fracture]
9. Abnormal Adult Shoulder [Dislocation]
10. Abnormal Adult Knee [Avulsion Fracture]
11. Abnormal Adult Elbow [Multiple Fracture]
12. Normal Adult Ankle
13. Normal Adult Scaphoid
14. Abnormal Paediatric Clavicle [Fracture]
15. Normal Paediatric Ankle
16. Normal Adult Fingers
17. Normal Paediatric Elbow
18. Abnormal Adult Foot [Multiple Fractures]
19. Normal Paediatric Foot
20. Abnormal Adult Foot [Fracture and Dislocation]
21. Abnormal Paediatric Finger [Fracture]
22. Normal Adult Elbow
23. Normal Adult Wrist
24. Normal Adult Knee
25. Abnormal Paediatric Ankle [Fracture]
Table 2a. Radiographers R1-7 individual performance.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Test Score (%)</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
<th>Cohens Kappa (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>36.5 (96%)</td>
<td>92.3 (72.4-92.3)</td>
<td>100 (78.4-100)</td>
<td>92.0 (50.6-92.0)</td>
</tr>
<tr>
<td>R2</td>
<td>29.5 (78%)</td>
<td>65.4 (43.5-72.7)</td>
<td>91.7 (67.9-99.6)</td>
<td>56.4 (11.2-71.4)</td>
</tr>
<tr>
<td>R3</td>
<td>32.25 (85%)</td>
<td>78 (55.2-85.6)</td>
<td>92 (69.2-99.6)</td>
<td>70 (24.5-85.1)</td>
</tr>
<tr>
<td>R4</td>
<td>30 (79%)</td>
<td>67.3 (45.3-74.6)</td>
<td>91.7 (67.9-99.6)</td>
<td>58.4 (13.1-7.34)</td>
</tr>
<tr>
<td>R5</td>
<td>34.5 (90%)</td>
<td>88.5 (67-95.7)</td>
<td>91.7 (68.4-99.5)</td>
<td>80 (35.4-95.1)</td>
</tr>
<tr>
<td>R6</td>
<td>34.5 (90%)</td>
<td>88.5 (67-95.7)</td>
<td>91.7 (68.4-99.5)</td>
<td>80 (35.4-95.1)</td>
</tr>
<tr>
<td>R7</td>
<td>32.25 (85%)</td>
<td>78 (55.2-85.6)</td>
<td>92 (69.2-99.6)</td>
<td>70 (24.5-85.1)</td>
</tr>
</tbody>
</table>

| Mean        | 32.7 (86%)     | 79.7 (71.8-87.5)    | 92.9 (90.6-95.1)    | 72.4 (63-81.7)       |
| Median      | 32.2           | 78                   | 91.7                | 70                   |
| Mode        | 32.2           | 78                   | 91.7                | 70                   |
| Standard Deviation | 2.5         | 10.6                 | 3.1                | 12.6                |

Table 2b. Nurses N1-19 individual performance.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Test Score (%)</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
<th>Cohens Kappa (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>33 (87%)</td>
<td>98.1 (77.0-100)</td>
<td>75.0 (52.2-77.1)</td>
<td>73.7 (29.5-77.8)</td>
</tr>
<tr>
<td>N2</td>
<td>33 (87%)</td>
<td>98.1 (77.0-100)</td>
<td>75.0 (52.2-77.1)</td>
<td>73.7 (29.5-77.8)</td>
</tr>
<tr>
<td>N3</td>
<td>37 (97%)</td>
<td>100 (79.1-100)</td>
<td>94.1 (74.0-94.1)</td>
<td>94.0 (53.0-94.0)</td>
</tr>
<tr>
<td>N4</td>
<td>34 (88%)</td>
<td>100 (81.1-100)</td>
<td>72.7 (48.6-72.7)</td>
<td>74.9 (30.6-74.9)</td>
</tr>
<tr>
<td>N5</td>
<td>35.5 (93%)</td>
<td>86.5 (66.1-86.5)</td>
<td>100 (77.9-100)</td>
<td>86.1 (43.7-86.1)</td>
</tr>
<tr>
<td>N6</td>
<td>36.25 (95%)</td>
<td>98.2 (79.4-100)</td>
<td>90.9 (67.0-93.2)</td>
<td>89.8 (46.8-93.9)</td>
</tr>
<tr>
<td>N7</td>
<td>33.25 (88%)</td>
<td>91.4 (69.7-99.4)</td>
<td>83.3 (59.8-92.0)</td>
<td>74.9 (29.5-91.6)</td>
</tr>
<tr>
<td>N8</td>
<td>23.5 (62%)</td>
<td>59.1 (33.2-80.9)</td>
<td>64.3 (44.8-81.4)</td>
<td>23.3 (0-62.1)</td>
</tr>
<tr>
<td>N9</td>
<td>23 (61%)</td>
<td>58.1 (31.9-80.7)</td>
<td>63.2 (43.4-80.2)</td>
<td>21.1 (0-60.3)</td>
</tr>
<tr>
<td>N10</td>
<td>21 (55%)</td>
<td>59 (30.9-83.7)</td>
<td>52.5 (34.5-68.3)</td>
<td>10.8 (0-49)</td>
</tr>
<tr>
<td>N11</td>
<td>26 (69%)</td>
<td>62.7 (39.8-78.8)</td>
<td>75.5 (51.6-92.2)</td>
<td>38.1 (0-70.8)</td>
</tr>
<tr>
<td>N12</td>
<td>32.5 (86%)</td>
<td>87 (62.1-98.1)</td>
<td>85.2 (64-94.7)</td>
<td>71.9 (26-92.5)</td>
</tr>
<tr>
<td>N13</td>
<td>23 (60%)</td>
<td>42.9 (23.3-54.5)</td>
<td>81.8 (57-96.6)</td>
<td>23.3 (0-48.2)</td>
</tr>
<tr>
<td>N14</td>
<td>30.5 (80%)</td>
<td>76.9 (54.4-89.1)</td>
<td>83.3 (59-96.5)</td>
<td>60.1 (12.4-85.3)</td>
</tr>
<tr>
<td>N15</td>
<td>27 (71%)</td>
<td>61.8 (40.7-73.6)</td>
<td>82.2 (56.4-96.6)</td>
<td>42.9 (0-68.4)</td>
</tr>
<tr>
<td>N16</td>
<td>28.5 (75%)</td>
<td>66 (41.3-80.7)</td>
<td>83 (61.2-96.1)</td>
<td>49.4 (02.5-77.4)</td>
</tr>
<tr>
<td>N17</td>
<td>33 (87%)</td>
<td>81 (63.6-82.5)</td>
<td>97.3 (67.7-100)</td>
<td>73.7 (29.5-77.8)</td>
</tr>
<tr>
<td>N18</td>
<td>26 (68%)</td>
<td>57.1 (36.5-68.7)</td>
<td>81.8 (55.6-96.5)</td>
<td>37.5 (0-62.8)</td>
</tr>
<tr>
<td>N19</td>
<td>28 (73%)</td>
<td>65.7 (49-71.3)</td>
<td>87.9 (54-99.4)</td>
<td>46.7 (2.5-61.7)</td>
</tr>
</tbody>
</table>

| Mean        | 29.6 (77.8%)   | 76.2 (67.9-84.4)    | 80.4 (75-85.7)      | 56.1 (44.6-67.5)     |
| Median      | 30.5           | 76.9                 | 82.2               | 60.1                 |
| Mode        | 33             | 98.1                 | 75                 | 73.7                 |
| Standard Deviation | 4.9         | 18.3                 | 11.9              | 25.5                 |
Table 2c. Paramedics P1-5 individual performance.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Test Score (%)</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
<th>Cohens Kappa (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>29.75 (78%)</td>
<td>80.8 (58.2-94.4)</td>
<td>75.0 (50.6-89.8)</td>
<td>55.9 (08.8-84.3)</td>
</tr>
<tr>
<td>P2</td>
<td>32.25 (85%)</td>
<td>94.2 (72.6-99.9)</td>
<td>75.0 (51.6-81.1)</td>
<td>69.7 (24.4-81.6)</td>
</tr>
<tr>
<td>P3</td>
<td>29.75 (78%)</td>
<td>80.8 (58.2-94.4)</td>
<td>75.0 (50.6-89.8)</td>
<td>55.9 (08.8-84.3)</td>
</tr>
<tr>
<td>P4</td>
<td>30 (79%)</td>
<td>74.5 (49.6-88.8)</td>
<td>83 (61-95.8)</td>
<td>57.7 (10.7-84.9)</td>
</tr>
<tr>
<td>P5</td>
<td>32 (84%)</td>
<td>71.4 (51.9-71.4)</td>
<td>100 (75.2-100)</td>
<td>68.8 (26.1-68.8)</td>
</tr>
<tr>
<td>Mean</td>
<td>30.7 (80.7%)</td>
<td>80.3 (72.6-87.9)</td>
<td>81.6 (72.1-91)</td>
<td>61.6 (55.4-67.7)</td>
</tr>
<tr>
<td>Median</td>
<td>30</td>
<td>80.8</td>
<td>75</td>
<td>57.7</td>
</tr>
<tr>
<td>Mode</td>
<td>29.7</td>
<td>80.8</td>
<td>75</td>
<td>55.9</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.2</td>
<td>8.7</td>
<td>10.8</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Table 2d. Physiotherapists PT1-2 individual performance.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Test Score (%)</th>
<th>Sensitivity (95%CI)</th>
<th>Specificity (95%CI)</th>
<th>Cohens Kappa (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT1</td>
<td>38 (100%)</td>
<td>100 (81.2-100)</td>
<td>100 (79.6-100)</td>
<td>100 (60.9-100)</td>
</tr>
<tr>
<td>PT2</td>
<td>31.25 (82%)</td>
<td>80.8 (58.4-92.7)</td>
<td>83.3 (59.1-96.3)</td>
<td>64.0 (17.4-88.9)</td>
</tr>
<tr>
<td>Mean</td>
<td>34.6 (91%)</td>
<td>90.4 (71.6-100)</td>
<td>91.6 (75.2-100)</td>
<td>82 (46.8-100)</td>
</tr>
<tr>
<td>Median</td>
<td>34.6</td>
<td>90.4</td>
<td>91.6</td>
<td>82</td>
</tr>
<tr>
<td>Mode</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.7</td>
<td>13.5</td>
<td>11.8</td>
<td>25.4</td>
</tr>
</tbody>
</table>

Table 3a. ANOVA of differences among group means in the study sample OSE Test score.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiographers</td>
<td>7</td>
<td>229.5</td>
<td>32.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Nurses</td>
<td>19</td>
<td>564</td>
<td>29.6</td>
<td>24.7</td>
</tr>
<tr>
<td>Paramedics</td>
<td>5</td>
<td>153.7</td>
<td>30.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Physiotherapists</td>
<td>2</td>
<td>69.2</td>
<td>34.6</td>
<td>22.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>80.5</td>
<td>3</td>
<td>26.8</td>
<td>1.5</td>
<td>0.23</td>
<td>2.9</td>
</tr>
<tr>
<td>Within Groups</td>
<td>512.9</td>
<td>29</td>
<td>17.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>593.4</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3b. ANOVA of differences among group means in the study sample sensitivity score.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiographers</td>
<td>7</td>
<td>558</td>
<td>79.7</td>
<td>112.9</td>
</tr>
<tr>
<td>Nurses</td>
<td>19</td>
<td>1449.6</td>
<td>76.2</td>
<td>335.9</td>
</tr>
<tr>
<td>Paramedics</td>
<td>5</td>
<td>401.7</td>
<td>80.3</td>
<td>76.6</td>
</tr>
<tr>
<td>Physiotherapists</td>
<td>2</td>
<td>180.8</td>
<td>90.4</td>
<td>184.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>402.8</td>
<td>3</td>
<td>134.2</td>
<td>0.5</td>
<td>0.65</td>
<td>2.9</td>
</tr>
<tr>
<td>Within Groups</td>
<td>7215.2</td>
<td>29</td>
<td>248.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7618</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3c. ANOVA of differences among group means in the study sample specificity score.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiographers</td>
<td>7</td>
<td>650.8</td>
<td>92.9</td>
<td>9.6</td>
</tr>
<tr>
<td>Nurses</td>
<td>19</td>
<td>1529</td>
<td>80.4</td>
<td>141.8</td>
</tr>
<tr>
<td>Paramedics</td>
<td>5</td>
<td>408</td>
<td>81.6</td>
<td>117.8</td>
</tr>
<tr>
<td>Physiotherapists</td>
<td>2</td>
<td>183.3</td>
<td>91.6</td>
<td>139.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>945.4</td>
<td>3</td>
<td>315.1</td>
<td>2.8</td>
<td>0.05</td>
<td>2.9</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3221.1</td>
<td>29</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4166.6</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Alternative Free-response Receiver Operating Characteristic (AFROC) curve plot of professional subgroup performance.
Figure 2. Literature search.

Keywords:
- Radiographer
- Nurse (Inc. Emergency Nurse Practitioner)
- Paramedic (Inc. Advanced Paramedic Emergency Practitioner)
- Physiotherapist (Inc. Advanced Musculoskeletal Physiotherapist)
- +/- X-ray, Red dot, Image Interpretation, Preliminary Clinical Evaluation

Citations:
- CENTRAL: n = 2
- CINHAL: n = 249
- Google Scholar: n = 18,300
- MEDLINE: n = 313,226
- PubMed: n = 41
- British NUR Index: n = 160
- BMJ Journals: n = 1,613
- Cambridge J. n = 5,292
- Internurse n = 3
- BMJ Journals: n = 57
- Europe PMC: n = 5
- SAGE Journals: n = 24
- Wiley: n = 73
- BioMed Central: n = 2

Citations excluded:
- After abstract review: n = 331,700
- Nurse: n = 348,042
- Paramedic: n = 9,193
- Physiotherapist: n = 18,63
- Keyword used out of context in study, method, result, discussion, case study, literature review, or duplicate study

Topic relevant studies:
- Radiographer: n = 118
- Nurse: n = 20
- Paramedic: n = 16
- Physiotherapist: n = 14

Studies excluded:
- Radiographer: Opinion piece n = 15, Qualitative study n = 8, Mammography n = 12, Cross-sectional n = 7, Dental n = 9, Missing data for 2x2 table n = 11, Archived Journal offline n = 1
- Nurse, Paramedic, Physiotherapist: Opinion piece n = 8, Qualitative study n = 7, X-rays / NG tubes n = 5, Triage of patients n = 6, Ultrasound n = 4, Missing data for 2x2 table n = 5

Eligible studies:
- Literature review: n = 41, n = 7
- Nurse: n = 20, n = 0
- Paramedic: n = 0
- Physiotherapist: n = 0
Figure 3 A. Radiographer literature analysis of sensitivity in musculoskeletal radiograph image interpretation (*pre-test results, ~mid-test results, ~post-test results, ^ follow-up test results).
Figure 3 B. Radiographer literature analysis of specificity in musculoskeletal radiograph image interpretation (*pre-test results, "mid-test results, ~post-test results, ^ follow-up test results).
Figure 3 C. Radiographer Summary Receiver Operating Characteristic (SROC) curve analysis.
Figure 4 A. Nurse literature analysis of sensitivity in musculoskeletal radiograph image interpretation (*pre-test results, “mid-test results, ~post-test results, ^ follow-up test results).

![Sensitivity Graph]

<table>
<thead>
<tr>
<th>Study</th>
<th>Sensitivity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freij et al 1996</td>
<td>0.94 (0.86 - 0.98)</td>
</tr>
<tr>
<td>Benger 2002</td>
<td>0.96 (0.90 - 0.99)</td>
</tr>
<tr>
<td>Derksen et al 2006*</td>
<td>0.87 (0.82 - 0.91)</td>
</tr>
<tr>
<td>Derksen et al 2006~</td>
<td>0.89 (0.84 - 0.93)</td>
</tr>
<tr>
<td>Coleman et al 2009</td>
<td>0.49 (0.40 - 0.68)</td>
</tr>
<tr>
<td>Piper et al 2009*</td>
<td>0.48 (0.42 - 0.54)</td>
</tr>
<tr>
<td>Piper et al 2009~</td>
<td>0.75 (0.70 - 0.80)</td>
</tr>
<tr>
<td>Lee et al 2014</td>
<td>0.91 (0.82 - 0.96)</td>
</tr>
<tr>
<td>Snaith et al 2014</td>
<td>0.92 (0.80 - 0.98)</td>
</tr>
</tbody>
</table>

Pooled Sensitivity = 0.75 (0.73 to 0.77)
Chi-square = 266.16; df = 8 (p = 0.0000)
Inconsistency (I-square) = 97.0%

Figure 4 B. Nurse literature analysis of specificity in musculoskeletal radiograph image interpretation (*pre-test results, “mid-test results, ~post-test results, ^ follow-up test results).

![Specificity Graph]

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freij et al 1996</td>
<td>0.93 (0.88 - 0.97)</td>
</tr>
<tr>
<td>Benger 2002</td>
<td>0.87 (0.82 - 0.91)</td>
</tr>
<tr>
<td>Derksen et al 2006*</td>
<td>0.87 (0.82 - 0.91)</td>
</tr>
<tr>
<td>Derksen et al 2006~</td>
<td>0.92 (0.88 - 0.95)</td>
</tr>
<tr>
<td>Coleman et al 2006</td>
<td>0.54 (0.45 - 0.63)</td>
</tr>
<tr>
<td>Piper et al 2009*</td>
<td>0.66 (0.57 - 0.74)</td>
</tr>
<tr>
<td>Piper et al 2009~</td>
<td>0.39 (0.31 - 0.48)</td>
</tr>
<tr>
<td>Lee et al 2014</td>
<td>0.85 (0.77 - 0.91)</td>
</tr>
<tr>
<td>Snaith et al 2014</td>
<td>0.99 (0.93 - 1.00)</td>
</tr>
</tbody>
</table>

Pooled Specificity = 0.80 (0.77 to 0.82)
Chi-square = 256.21; df = 8 (p = 0.0000)
Inconsistency (I-square) = 96.9%
Figure 4 C. Nurse Summary Receiver Operating Characteristic (SROC) curve analysis.

AUC = 0.9125
SE(AUC) = 0.0456