Title

Agreement between expert thoracic radiologists and the chest radiograph reports provided by consultant radiologists and reporting radiographers in clinical practice: review of a single clinical site

Shortened Title

Agreement in chest radiograph reporting

Key Words

Clinical Competence; Radiography, thoracic; Radiographer Reporting; Observer Performance

Abstract

Introduction: To compare the clinical chest radiograph (CXR) reports provided by consultant radiologists and reporting radiographers with expert thoracic radiologists.

Methods: Adult CXRs (n=193) from a single site were included; 83% randomly selected from CXRs performed over one year, and 17% selected from the discrepancy meeting. Chest radiographs were independently interpreted by two expert thoracic radiologists (CTR1/2).Clinical history, previous and follow-up imaging was available, but not the original clinical report. Two arbiters compared expert and clinical reports independently. Kappa (K), Chi Square (χ^2) and McNemar tests were performed to determine inter-observer agreement.

Results: CTR1 interpreted 187 (97%) and CTR2 186 (96%) CXRs, with 180 CXRs interpreted by both experts. Radiologists and radiographers provided 93 and 87 of the original clinical reports respectively. Consensus between both expert thoracic radiologists and the radiographer clinical report was 70 (CTR1;K=0.59) and 70 (CTR2; K=0.62), and comparable to agreement between expert thoracic radiologists and the radiologist clinical report (CTR1=76,K=0.60; CTR2=75, K=0.62). Expert thoracic radiologists agreed in 131 cases (K=0.48). There was no difference in agreement between either expert thoracic radiologist, when the clinical report was provided by radiographers or radiologists (CTR1 χ=0.056, p=0.813; CTR2 χ=0.014, p=0.906), or when stratified by inter-expert agreement; radiographer McNemar p=0.629 and radiologist p=0.701.

Conclusion: Even when weighted with chest radiographs reviewed at discrepancy meetings, content of CXR reports from trained radiographers indistinguishable from content of reports issued by radiologists and expert thoracic radiologists.

Introduction

Clinical imaging is fundamental to modern medicine and worldwide there has been a sustained increase in demand for all radiological investigations.[1-4] The chest radiograph is integrated into many patient pathways, and is the most frequent radiology examination in England with approximately 6.7 million performed in 2015-16.[5] The rise in demand for imaging has not been matched with increased diagnostic capacity.[6] The Royal College of Radiologists (RCR) has found that more than 230,000 imaging examinations across all modalities are waiting more than 30 days to be reported,[7] of which 170,00 are plain imaging investigations. Further, the volume of non-radiology clinicians who interpret their own examinations is increasing.[8] Clinical reporting by trained radiographers is established in the UK, and used by many departments to meet rising demand.[8-10] Recent work by Milner *et al.* suggests that this is concentrated on musculoskeletal radiograph reporting (255 of 259 respondents, 98.5%) with only 39 individual radiographers (15.1%) indicating that they currently report CXRs in practice.[11] There is definitive evidence that reporting radiographers interpret skeletal radiographs with high accuracy.[12 13] Evidence regarding chest radiograph reporting by trained radiographers is more limited.[14 15] Several earlier studies have examined the performance of radiographers in identifying normal and abnormal CXRs with promising results.[16-18] However, the radiographers had not completed accredited postgraduate CXR reporting training.

The aim of this study was to compare the content of clinical CXR reports, which had been provided by consultant radiologists and reporting radiographers with expert thoracic radiologists during the construction of a standard reference bank of CXRs.

Methods

Design, setting and ethical approval

This retrospective study, mirroring a case-control design, was conducted in an acute district general hospital that performed approximately 20,000 hospital based CXRs per year. The purpose of the study was to establish an image bank with a robust reference standard diagnosis to examine diagnostic accuracy. Ethical approval was obtained from xxxx research ethics committee prior to the study commencing.

Case selection

A sample of anonymised adult (>16 years) CXRs performed over a consecutive twelve month (April 2010 – March 2011) period at a single London acute district general hospital was used. A total of 106 CXRs (normal n=53, 50%) were required for the image bank. An estimated 176 cases were required to be reviewed to generate a test bank of 106 CXRs with agreed reports assuming 61% agreement between experts.[19] Sim and Wright estimate that for a disease prevalence of 50% and Kappa (K) of 0.5 a total of 43 cases would give 90% power to detect a significant difference at p=0.05.[20] All CXRs were acquired using computed radiography (CR) or direct digital radiography (DR) systems. The time frame was chosen to enable all clinical or radiological follow up to be performed in order to provide maximum clinical information to the expert thoracic radiologists so that a robust reference standard diagnosis could be formed. Comparison of case mix and the use of follow-up images and CT scans as not changed substantially between data collection (2010) and publication (2018; local audit data). Inclusion criteria were patients referred by a hospital based clinician; emergency department, outpatient and inpatient examinations. Chest radiographs of patients under 16 years of age, referrals from general practice and multiple CXRs from a single patient were excluded. Referrals from general practice were excluded due to logistical difficulties in obtaining case notes for review as part of the reference standard diagnosis.

Cases were also drawn from the monthly radiology discrepancy meeting (32, 17%). Reports were provided by both consultant radiologists (n=8) and reporting radiographers (n=2). Selection of difficult cases from the discrepancy meeting ensured that the resultant image bank included a range of subtle cases (difficult normal and abnormal) to ensure that it would be discriminatory for small differences in observer performance for the study of diagnostic accuracy.

To ensure that a representative sample of normal and abnormal cases and a range of pathologies were included in the image bank that was to be formed, cases were stratified for a normal:abnormal ratio of 1:1 and, for abnormal cases, a disease category (infection: cardiopulmonary: malignancy: other) ratio of 3:3:1:3. Examples of cardiopulmonary pathology included pulmonary oedema, congestive cardiac failure and pericardial effusion. The percentage of cases within each broad disease category was based on the proportions found at audit of most frequent discharge diagnoses associated with CXRs performed at the study site. These proportions were matched against national disease datasets, and found to be similar.[21] Stratification of the cases (normal and for each disease category) was performed, based on the clinical report provided by the reporting practitioner at time of clinical interpretation. Stratification of CXRs was consistent for reporting radiographer and consultant radiologist reports.

Reporting

Chest radiograph reports were provided by reporting radiographers (post accredited education experience 1 and 3 years) and consultant radiologists (n=8; experience 1-20 years post-FRCR). Two consultant radiologists with a subspecialist interest in thoracic imaging (consultant thoracic radiologists; CTR1/CTR2) independently interpreted the CXRs, blinded to the clinical report. All pertinent imaging (previous and follow up CXRs, cross-sectional imaging) were available, and patient demographics and clinical history provided. Best practice in reporting is to review previous imaging when available. Additional imaging was made available to the thoracic radiologists so that a robust reference standard diagnosis could be obtained for the image bank. Features to be considered

normal (incidental findings) and significant (abnormal) were agreed in advance, based on the work of Robinson *et al*.[19] The expert thoracic radiologists indicated if the CXR was normal or abnormal, and for all abnormal cases, identified, localised and provided a diagnosis for all abnormalities present.

Report comparison

Two independent arbiters compared both the interpretations of the expert thoracic radiologists and each thoracic radiologist interpretation with the clinical report. Both arbiters had experience in comparing radiology reports for agreement in academic practice as part of clinical reporting assessment. Reports were determined to be in concordance only when both independent arbiters agreed that all abnormalities were identified and localised. Arbiters were blinded to the source of the report and did not have access to the images, patient demographics or clinical history.

Statistical analysis

Inter-observer agreement, between thoracic radiologists and between each thoracic radiologist and the clinical report, was determined using Kappa (K) statistic. Agreement was categorised according to Landis and Koch.[22] Moderate ($0.41<K<0.60$) agreement corresponds to 15-35% of data that is reliable and substantial ($0.61<K<0.80$) translates to 35-63% reliability.[23] Chi square and McNemar's test were used to examine the proportion of cases where the thoracic radiologist interpretations were in concordance with the clinical report provided by consultant radiologists and reporting radiographers. For this measure agreement was exact, missing no pathologies and not adding any other findings, rather than an agreement at the level of a clinical diagnosis. Statistical difference between observers was determined by examination of the 95% confidence intervals for Kappa; overlapping indicates no statistical significant difference for all analyses and sub analyses.[24]

Results

A summary of cases included and reviewed by expert thoracic radiologists is presented in Figure 1.

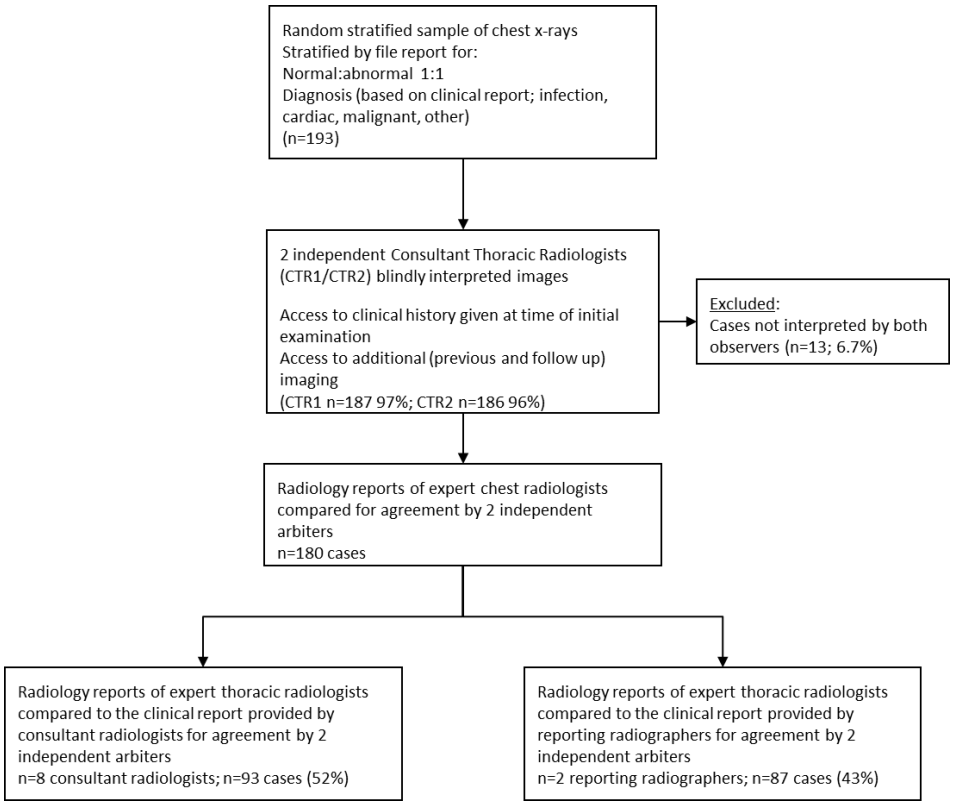

Figure 1. Summary of case inclusion and expert thoracic radiologist review

Case inclusion and characteristics

One hundred and six CXRs were required for the image bank (n=53 (50%) normal; n=18 (17%) infection; n=14 (13%) cardiopulmonary; n=7 (7%) malignancy; n=14 (13%) other). A total of 193 cases were reviewed by the expert thoracic radiologists to obtain the required image bank (n=106 CXRs). Each thoracic radiologist interpreted 187 (97%; CTR1) and 186 (96%; CTR2) examinations respectively. Thirteen cases that received a single thoracic radiologist report (7%) were excluded from analysis. The small number of cases not interpreted by each expert (CTR1 6 cases; CTR2 7 cases) occurred due to expert oversight (no image available for review, data form not completed). There was no pattern to the cases only interpreted by a single expert radiologist and the cases are summarised in Table 1. Of the 180 cases included, 52% (93 of 180) had the clinical report provided

by a consultant radiologist and 48% (87 of 180) by a reporting radiographer. Table 2 shows the prior and follow-up imaging relative to the cases that were included in the study.

Expert Consultant Thoracic Radiologist	Number of Cases
Thoracic Radiologist (CTR) 1	Infection n=2 Cardiopulmonary n=1 Other n=3
Thoracic Radiologist (CTR) 2	Normal n=3 Cardiopulmonary n=1 Malignant n=2 Other n=1

Table 1. Summary of cases that only received a single cardiothoracic radiologist report

	Number of Cases
Previous CXR	35 (19%)
Follow Up CXR	30 (17%)
Both Previous and Follow Up CXR	83 (46%)
Neither Previous nor Follow Up CXR	32 (18%)

Table 2. Proportion of cases with previous and/or follow up imaging

Agreement

For dichotomous normal/abnormal decisions, Table 3 shows moderate (0.41<K<0.60) or substantial (0.61<K<0.80) agreement between expert radiologists and the clinical reports of consultant radiologists and reporting radiographers using the classification system of Landis & Koch.[22] The statistical significance ($p<0.05$) of Kappa values for most analyses indicates that the estimated agreement between observers (CTR-RR, CTR-CR, inter-CTR) was probably not due to chance.[23] Access to additional imaging did not appear to influence agreement between thoracic radiologists and reports provided by reporting radiographers or consultant radiologists as evidenced by overlapping 95% confidence intervals for Kappa between all observers for all sub-analyses.[24] The CXRs with follow-up images did not show greater agreement than those with only a single radiograph.

	Agree Normal (Expert-Clinical)	Agree Abnormal (Expert-Clinical)	Total Cases	Kappa (95% CI)	Kappa p value
With Previous Imaging					
Reporting Radiographer					
Expert CTR1	9	6	21	0.46 (0.15 – 0.78)	0.012*
Expert CTR2	13	6	21	0.79 (0.51 – 1)	<0.0001*
Consultant Radiologist					
Expert CTR1	3	10	14	0.81 (0.45 – 1)	0.001*
Expert CTR2	4	8	14	0.70 (0.32 – 1)	0.006*
Thoracic Radiologists	11	16	35	0.60 (0.35 – 0.850)	<0.0001*
With Follow Up Imaging					
Reporting Radiographer					
Expert CTR1	2	9	12	0.75 (0.29 – 1)	0.007*
Expert CTR2	2	7	12	0.44 (0 – 0.9)	0.067
Consultant Radiologist					
Expert CTR1	4	7	18	0.23 (0 – 0.64)	0.196
Expert CTR2	8	5	18	0.43 (0 – 0.86)	0.066
Thoracic Radiologists	9	14	30	0.50 (0.24 – 0.77)	0.001*
Both Previous and Follow Up Imaging					
Reporting Radiographer					
Expert CTR1	7	27	40	0.60 (0.31 – 0.89)	<0.0001*
Expert CTR2	9	21	40	0.47 (0.21 – 0.73)	0.001*
Consultant Radiologist					
Expert CTR1	5	31	43	0.50 (0.18 – 0.81)	<0.0001*
Expert CTR2	11	24	43	0.61 (0.37 – 0.84)	<0.0001*
Thoracic Radiologists	15	45	83	0.41 (0.23 – 0.58)	<0.0001*
Neither					
Reporting Radiographer					
Expert CTR1	7	3	14	0.43 (0 – 0.83)	0.051
Expert CTR2	11	1	14	0.44 (0 – 1)	0.047*
Consultant Radiologist					
Expert CTR1	11	5	18	0.75 (0.44 – 1)	0.001*
Expert CTR2	12	3	18	0.56 (0.11 – 1)	0.017*
Thoracic Radiologists	17	4	32	0.25 (0 – 0.53)	0.075
Combined					
Reporting Radiographer					
Expert CTR1	25	45	87	0.59 (0.42 – 0.76)	<0.0001*
Expert CTR2	35	35	87	0.62 (0.43 – 0.78)	<0.0001*
Consultant Radiologist					
Expert CTR1	23	53	93	0.60 (0.44 – 0.77)	<0.0001*
Expert CTR2	35	40	93	0.62 (0.46 – 0.77)	<0.0001*
Thoracic Radiologists	52	79	180	0.48 (0.36 – 0.59)	<0.0001*

Table 3. Influence of additional imaging on normal-abnormal agreement between expert radiologists and the clinical report
*=statistically significant result ($p<0.05$)

Report concordance

Table 4 shows comparable report concordance between reports provided by expert thoracic radiologists, consultant radiologists and reporting radiographers (CTR1 Chi square χ^2=0.056, p=0.813; CTR2 χ^2=0.014, p=0.906). Both arbiters agreed that the abnormalities described by expert thoracic radiologist (CTR1) were included on the list of insignificant findings according to study protocol for a single instance (stable post-operative change), and thus should have been considered normal.

	Disagree	Agree	Total Cases
Reporting Radiographer			
Expert CTR1	38 (44%)	49 (56%)	87
Expert CTR2	32 (37%)	55 (63%)	87
Consultant Radiologist			
Expert CTR1	39 (42%)	54 (58%)	93
Expert CTR2	35 (38%)	58 (62%)	93

Table 4. Expert thoracic radiologist report concordance with the clinical report

Concordance between the thoracic radiologists and the clinical reports was stratified by inter-expert agreement, to account for variability between the expert radiologists. Both experts agreed with similar proportions of reports provided by reporting radiographers (CTR1-RR 48%; CTR2-RR 47%) and consultant radiologists (CTR1-CR 42%; CTR2-CR 44%) as well as cases where there was disagreement between both experts and the clinical reports (Table 5). When variability between expert thoracic radiologists was accounted for, no significant difference in report concordance was found between consultant radiologist (McNemar, p=0.701) and reporting radiographer (McNemar, p=0.629) clinical reports.

	Inter-Expert Agreement (CTR1-CTR2)		
	Disagree	Agree	Total Cases
Reporting Radiographer			
Expert CTR1	28 (32%)	42 (48%)	87
Expert CTR2	21 (24%)	41 (47%)	87
Consultant Radiologist			
Expert CTR1	27 (29%)	39 (42%)	93
Expert CTR2	25 (27%)	41 (44%)	93

Table 5. Complete report concordance between expert radiologists (CTR1 & CTR2) and the clinical reports provided by consultant radiologists and reporting radiographers

Discussion

In this study, we found that agreement among trained reporting radiographers, radiologists and expert thoracic radiologists did not differ, even when a high proportion of difficult CXRs was included in the image bank. Agreement was especially high for normal CXRs. The stringent nature of the agreement criteria gave lower results than in the current literature. However, no difference in reporting agreement could be demonstrated among the three groups.

Chest radiograph interpretation accuracy

The significant body of evidence that examines the performance of observers when interpreting CXRs confirms that considerable inter-observer variation exists.[25-29] Accordingly, this was taken into account for data analyses in the current study when comparing agreement between expert thoracic radiologists and the clinical reports of consultant radiologists and reporting radiographers.

The current study found the expert chest radiologists agreed in 52 (of 180 total cases, 29%) of the normal cases and 54 (of 180 cases, 30%) of abnormal cases, with moderate agreement using the classification of Landis & Koch.[22] Both the work of Robinson *et al*., who found three consultant radiologists agreed in only 61 of 100 CXRs (K=0.50),[19] and Tudor *et al*. where agreement was K=0.58 (95%CI .49-0.67)[30] when a bank of 50 radiographs were interpreted with clinical history by five

radiologists , found slightly higher agreement than the inter-expert radiologist agreement in the current study (K=0.48). The larger sample size (n=180) and use of reports produced in clinical practice in the current study may partially explain the lower reported agreement. The use of strict criteria when determining report agreement, as required for the reference standard diagnosis, could also contribute to the lower agreement in the current study. Importantly, agreement between expert thoracic radiologists and the clinical reports of consultant radiologists and reporting radiographers were comparable and in line with previous research that determined agreement in CXR reporting between consultant radiologists. Equivalent performance of reporting radiographers and consultant radiologists is required.[31] The results of the current study suggest that CXR reporting performance, and by inference patient safety, is maintained with radiographer CXR in clinical practice.

Radiographer chest radiograph reporting

This is the first study to examine trained radiographer reporting of CXRs appropriate to secondary care and difficult CXRs. Other studies have noted that radiographers can distinguish between normal and abnormal images,[17 18] and can detect lung cancer in screening programs.[16] Fifty-one Nigerian radiographers interpreted a bank (n=50 CXRs, 27 abnormal) with lower sensitivity (76.9%; 95% CI 0.658 – 0.864) and specificity (79.8; 95% CI 0.658 – 0.864) to previous work, possibly due to the different educational and postgraduate experience profiles between the USA, UK and Nigeria.[32]

Only three previous studies that utilised trained reporting radiographers to report CXRs in clinical practice have been published. Brealey included a small number of CXRs (112 of 28,900) in a meta-analysis, which examined plain radiographic reporting performance, but accuracy was grouped for all body areas rather than the chest *per se*. Robinson *et al*. incorporated a small number of CXRs from the emergency department as part of a larger study, and agreement with a radiologist report found in 52 normal and 31 abnormal cases (83 of 112, 74%).[33] Both Brealey and Robinson failed to segregate CXRs from other examinations. The current results are lower than that reported by

Woznitza *et al*.[15] In their audit of clinical practice, 99 cases initially reported by a single reporting radiographer were reviewed independently by three consultant radiologists. The authors reported radiographer – radiologist agreement of 92%, 96% and 96%, and Kappa of 0.83, 0.91 and 0.91 respectively. These results may have been inflated due to incorporation and verification bias as the consultant radiologists did not perform their interpretations blinded to the radiographer report.[34] These limitations were avoided in the current study as the thoracic radiologists were blinded to the clinical report. The arbiters did not have access to the radiographs when comparing the reports for agreement, and the reports were blinded to origin. The results of the current study, which included qualified reporting radiographers as participants and was free from the biases identified above due to the robust methodology employed, found comparable agreement for reporting radiographers and consultant radiologists.

In the current study, absolute agreement between all findings was required for the case to be concordant; agreement was assessed between report content and not clinical report outcome or influence on patient treatment. No distinction was made on the clinical significance or context of the finding, for example a small unilateral pleural effusion in a patient with other radiological signs and clinical features of congestive cardiac failure. These strict criteria were required as the expert thoracic radiologist findings were to be used as the reference standard diagnosis for a diagnostic accuracy image bank. This may have resulted in lower inter-observer agreement, between expert radiologist and between the clinical reports provided by consultant radiologists and reporting radiographers. Crucially, CXR reporting agreement was lower only in absolute values when compared to previous literature but was not lower in relative terms between the three groups of observers (reporting radiographers, consultant radiologists, expert thoracic radiologists) in the current study. The criteria used for report concordance was applied consistently across all comparisons. Thus, current results suggest that CXR reporting between reporting radiologists and consultant radiologists is indistinguishable.

Limitations

Agreement between the clinical reports of consultant radiologists and reporting radiographers with expert thoracic radiologist review was the outcome used in the current study. The relative diagnostic accuracy of consultant radiologists and reporting radiographers was not assessed. Cases where all participants agreed on the findings could have been incorrect, although this was mitigated as the expert thoracic radiologists had access to follow up imaging (CXRs and CT) performed over 12-24 months after the index CXR.

Expert thoracic radiologist access to follow up imaging might have explained the lower agreement with the clinical reports (reporting radiographer and consultant radiologist) but this did not appear to be the case in this study. For example, expert thoracic radiologist access to CT could have confirmed or refuted mediastinal lymphadenopathy. However, even with this additional information, the agreement of reporting radiographers and radiologists with the expert thoracic radiologists did not differ.

The current study only included a single clinical site and a small number of practitioners, two expert thoracic radiologists, eight consultant radiologists and two reporting radiographers. This small sample limits its generalisability. However, the hospital was not a teaching hospital but rather a district general hospital. The exclusion of CXRs referred from primary care may have influenced the case mix included in the study, both in terms of diagnoses and disease severity. Lack of access to case notes for patients referred from primary care precluded inclusion of the CXRs in the diagnostic accuracy image bank. Notably, agreement was greater for normal chest radiographs, which constitute a higher proportion of the workload from primary compared to secondary care. Future research that incorporates a larger number of observers from multiple clinical sites is needed.

<u>Conclusion</u>

There is a general paucity of evidence that has examined the performance of radiographers interpreting CXRs in clinical practice. This study has found comparable levels of normal – abnormal agreement between clinical reports provided by consultant radiologists and reporting radiographers at a single clinical site. Comparable report concordance was also found between the expert chest consultant radiologists and the clinical reports of consultant radiologists. Even when weighted with chest radiographs reviewed at discrepancy meetings, reporting radiographers were indistinguishable from radiologists in their CXR reports.

Funding

The study was part funded by a xxxx research grant.

Conflict of interest

The authors have no conflicts of interest to declare.

References

1. Radiologists Royal College of. Clinical radiology UK workforce census 2015 report. 2016 London: Royal College of Radiologists
2. Team Radiology Service Improvement. Radiology: A National Framework for Service Improvement. 2003: Department of Health
3. Development Organisation for Economic Co-operation and. Health at a Glance: Europe 2014. 2014: OECD
4. Information Canadian Institute for Health. Health Care in Canada: A Focus on Wait Times. 2012 Ottowa, Canada: Canadian Institute for Health Information
5. England NHS. Diagnostic Imaging Dataset Statistical Release. London: Department of Health, 2016.
6. Delivery 2020. Horizon Scanning: An evaluation of imaging capacity across the NHS in England. 2015: Cancer Research UK
7. Radiologists Royal College of. Our patients are still waiting …. 2016 London: Royal College of Radiologists
8. Network NHS Benchmarking. Radiology Benchmarking 2016 National report. 2016 London: NHS Benchmarking Network
9. Woznitza N., Piper K., Rowe S. *et al.* Optimizing patient care in radiology through team-working: A case study from the United Kingdom. Radiography 2014;**20**(3):258-263.
10. Snaith B., Milner R. C., Harris M. A. Beyond image interpretation: Capturing the impact of radiographer advanced practice through activity diaries. Radiography 2016;**22**(4):e233-e238.
11. Milner R. C., Culpan G., Snaith B. Radiographer reporting in the UK: is the current scope of practice limiting plain-film reporting capacity? Br J Radiol 2016;**89**(1065):20160228.
12. Brealey S., Scally A., Hahn S. *et al.* Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. Clin Radiol 2005;**60**(2):232-241.
13. Piper K., Paterson A., Godfrey R. Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases. Radiography 2005;**11**(1):27-34.
14. Piper K., Cox S., Paterson A. *et al.* Chest reporting by radiographers: Findings of an accredited postgraduate programme. Radiography 2014;**20**(2):94-99.
15. Woznitza N., Piper K., Burke S. *et al.* Adult chest radiograph reporting by radiographers: Preliminary data from an in-house audit programme. Radiography 2014;**20**(3):223-229.
16. Flehinger B. J., Melamed M. R., Heelan R. T. *et al.* Accuracy of chest film screening by technologists in the New York early lung cancer detection program. AJR Am J Roentgenol 1978;**131**(4):593-597.
17. Sheft D. J., Jones M. D., Brown R. F. *et al.* Screening of chest roentgenograms by advanced roentgen technologists. Radiology 1970;**94**(2):427-429.
18. Sonnex E. P., Tasker A. D., Coulden R. A. The role of preliminary interpretation of chest radiographs by radiographers in the management of acute medical problems within a cardiothoracic centre. British Journal of Radiology 2001;**74**(879):230-233.
19. Robinson P., Wilson D., Coral A. *et al.* Variation between experienced observers in the interpretation of accident and emergency radiographs. British Journal of Radiology 1999;**72**(856):323-330.
20. Sim J., Wright C. C. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Physical Therapy 2005;**85**(3):257-268.
21. Centre Health and Social Care Information. Hospital Episode Statistics for England. Inpatient statistics, 2011-12. 2012: Health and Social Care Information Centre
22. Landis J. R., Koch G. G. The measurement of observer agreement for categorical data. Biometrics 1977;**33**(1):159-174.
23. McHugh M. L. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;**22**(3):276-282.

24. Gwet Kilem L. Testing the Difference of Correlated Agreement Coefficients for Statistical Significance. Educ Psychol Meas 2015;**76**(4):609-637.
25. Chen Y., James J. J., Dong L. *et al.* Measuring performance in the interpretation of chest radiographs: a pilot study. Clin Radiol 2016.
26. Potchen E. J., Cooper T. G., Sierra A. E. *et al.* Measuring performance in chest radiography. Radiology 2000;**217**(2):456-459.
27. Quekel L. G., Kessels A. G., Goei R. *et al.* Miss rate of lung cancer on the chest radiograph in clinical practice. Chest 1999;**115**(3):720-724.
28. Shah P. K., Austin J. H., White C. S. *et al.* Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect. Radiology 2003;**226**(1):235-241.
29. Shang H., Jiang Y., Li F. *et al.* ROC Curve for Extremely Subtle Lung Nodules on Chest Radiographs Confirmed by CT Scan. Acad Radiol 2016;**23**(3):297-303.
30. Tudor G. R., Finlay D. B., Taub N. An assessment of inter-observer agreement and accuracy when reporting plain radiographs. Clin Radiol 1997;**52**(3):235-238.
31. Radiologists Royal College of, Radiographers Society and College of. Team working in clinical imaging. 2012 London: Royal College of Radiologists and the Society and College of Radiographers
32. Ekpo E. U., Egbe N. O., Akpan B. E. Radiographers' performance in chest X-ray interpretation: the Nigerian experience. Br J Radiol 2015;**88**(1051):20150023.
33. Robinson P. J. Plain film reporting by radiographers--a feasibility study. Br J Radiol 1996;**69**(828):1171-1174.
34. Brealey S., Scally A. J. Bias in plain film reading performance studies. Br J Radiol 2001;**74**(880):307-316.