CoRIPS Research Award 086
Nick Woznitza
Awarded £9720
Establishing the diagnostic accuracy of radiographer chest x-ray reports and their influence on clinicians' clinical reasoning: A comparison with consultant radiologists

Lay Summary
It is essential that x- rays reports are accurate to help provide a correct diagnosis. It is not known whether clinicians view an x-ray report differently depending on whether it has been issued by a radiographer (non-medical practitioner) or by a consultant radiologist (medical practitioner), and if the report source influences their clinical diagnosis and reasoning.

This study will compare the accuracy of radiographers' and radiologists' chest x-ray reports, and investigate whether the origin of x-ray reports - radiographer or radiologist - influences clinicians' reasoning and their decisions for regarding patient diagnosis.

Description of the project:

Principal Aim
To demonstrate the accuracy of chest x-ray (CXR) reporting and whether the influence of a CXR report on clinicians' diagnosis and treatment decisions in affected by the source of the report.

Primary Research Question
In the context of reports produced for hospital based adult patients, are the chest x-ray reports produced by reporting radiographers equivalent in accuracy and influence on clinicians clinical reasoning to those produced by consultant radiologists?

Secondary Research Questions
Phase 1: What is the accuracy of reporting radiographer (RR) and consultant radiologist (CR) interpretations of adult chest x-rays (CXRs) from hospital based patients in a simulated environment using two independent chest CRs in agreement as the reference standard?

Phase 2: Is there any clinically significant difference between the influence that radiographer and radiologist CXR reports on hospital based patients have on clinicians' clinical reasoning and decision-making in a simulated environment?

Outcomes
Phase 1 Diagnostic Accuracy: Relative accuracy of RR & CR CXR interpretation, using sensitivity, specificity and area under the curve (AUC) of receiver operator characteristic (ROC) curves.

Phase 2 Influence on Diagnosis: Difference in proportion of RR and CR reports that have an influence on the clinical reasoning, using changes in diagnostic confidence.

Review of literature and identification of current knowledge gaps
Government targets, person-centred care, developing technology and an aging population have resulted in an unprecedented rise in imaging workload (1-4). In response to these increasing demands trained radiographers now undertake image interpretation (5,6). There is limited high quality research underpinning the benefit of many radiological investigations despite the considerable cost burden that many of these procedures bear (7). It is also not clear whether a CR or RR-derived report influences clinical decisions.

Image interpretation is a subjective task (8), and studies demonstrate significant variation in x-ray interpretation between radiologists (9-17). While the evidence for the accuracy of trained radiographers reporting skeletal radiographs is definitive (1), there has been little work comparing the accuracy of practising RRs to CRs in CXR interpretation (18,19).

Radiology investigations are frequently used by clinicians to reduce uncertainty by providing additional information (20,21). Studies have examined the role of other imaging modalities (22-28), however there is limited work examining the impact of chest x-ray reports. The major work examining the influence of x-rays, conducted 35 years ago in the United States, confirmed that radiology reports influence clinicians' diagnostic thinking. In this study CXR reports produced by CRs were found to lead to a new most important diagnosis in 50% of cases in an A&E setting (29). Only one study examined the impact that incorrect radiographer skeletal reports had on patient management, reporting that patient care was more negatively influenced by incorrect CR reports (30). There appears to be no work examining the influence of RR CXR reports on clinicians' clinical decision-making.

If RRs are shown to interpret CXR with comparable accuracy to CRs, and there is no clinically significant difference in the influence that these reports have on clinicians' clinical decision-making, RRs could provide an additional reporting resource to the NHS in an efficient and effective way. This could increase the volume and timeliness of reporting, enable streamlined patient pathways and improved patient care, while maximising the limited resources available.

This research will form part of the work submitted for the award of PhD.

Methodology & Methods

i) Methodology

A quasi-experimental approach will be used for both phases of the research programme. The diagnostic accuracy study (Phase 1) has been constructed using the Standards for Reporting Diagnostic accuracy (STARD) framework (31), incorporating the suggestions of Brealey & Scally (8). Methodological issues identified in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)(32) statement and the Consolidated Standards of Reporting Trials (CONSORT) extensions for non-inferiority(33) and cluster designs(34) have been integrated in the Phase 2 study to provide a robust research protocol.

ii) Methods

Phase 1 – Diagnostic Accuracy

Sampling

Ten CR, ten RR participants and two arbiters will be recruited through convenience sampling. A sample of from adult patients in which a CXR was performed will be retrospectively selected from a London acute DGH, stratified for a normal/abnormal ratio 1:1 and disease category (infection/cardiac/malignancy/other), based on proportions from a recent audit of most frequent discharge diagnoses associated with a CXR. CR and RR participants not appropriately qualified or not currently in practice will be excluded, as will CXR from paediatric patients and those referred from general practice. Any case in which there is disagreement against the reference standard will be excluded from the image bank.

Sample Size

Adopting an alternate free-response methodology ROC curves will be used (14,35). To detect a 10% difference in the area under the curve (AUC) and assuming 10 observers in each group and a normal/abnormal CXR ratio of 1:1, 101 cases are required to produce an 80% power of sample with a 5% possibility of a Type I error.

Data Collection Methods

A stratified selection of CXRs will be anonymised, coded and reported to be normal or abnormal by two expert chest CRs with supporting free-text report. A list of incidental findings to be considered normal will be agreed in advance. A robust reference standard will be created by these expert CRs. RRs (n=10) and CRs (n=10) will interpret this bank of CXRs independently, blinded to the reference standard diagnosis, state their confidence in the case is normal or abnormal on a 5 point scale and produce a free-text comment, standard for assessments of CXR interpretation accuracy(12,15). Participants will have access to previous CXRs, a list of `normal' incidental findings but no other investigation or report. A proportion of cases will be included in 2 viewing sessions to establish intra-observer variability. Two independent arbiters, blinded to the source of report, will then compare the RR and CR reports for agreement with the reference standard. All data will be anonymised and coded.

Analysis

Participant RR and CR sensitivity and specificity will be calculated. ROC curves will be constructed using the alternate free-response receiver operating characteristic methodology which accounts for multiple abnormalities in a single case (10, 13, 35). Participant CRs performance will be compared with the RR results (10,13, 35). Paired t-tests will be used to compare the ROC AUC for individual observers (36), with any difference greater than 10% deemed clinically significant. Intra-observer agreement, for the expert chest CR, participant RR & CR and arbiters will be determined by duplicating 10% of cases over a minimum of two sessions, and Kappa used to identify any trends (16, 36). Service User feedback will be integrated into the data analysis to focus the results on patient-important outcomes.

Phase 2 – Influence of CXR Reports

Sampling

The CXR cases from the Phase 1 diagnostic accuracy arm, and the reports produced by the reporting practitioners will be used. 24 participant physicians' will be recruited, using purposive sampling, from the same hospital from which the CXR cases were selected, with all physicians approached to participate. For the purposes of this study, the consultant grade will comprise of staff that hold a substantive consultant post within the hospital and specialist registration with the General Medical Council. Specialist registrars will be qualified medical practitioners on a training programme, while junior medical staff will be qualified medical practitioners who have not yet begun specialist training. Any cases in which the case notes are not available will be excluded and examined for trends.

Sample Size

Assuming that 50% of CR CXR reports will produce a new most important diagnosis(29), and utilising a pre-defined clinically insignificant difference of <10%, a non-inferiority study will require 310 cases in each arm to have an 80% power of sample with a 5% possibility of a Type I error (37). Each reporting practitioner is considered the source of a cluster of data. To account for the lack of independence (34, 38), a revised sample size of n=914 is required.

Data Collection Methods

Pre and post-CXR report proformas, based on work previously conducted (24,29,30), have been designed, and will be piloted to ensure reliability and validity prior to commencement of the study. They contain patient demographics, referral source and case summary. Eight clinicians will be recruited at each level, from a range of specialities and be randomly assigned to independently review 50% of cases. Eight consultant physicians represent 25% of the full time consultant physician posts from the site of recruitment. The outcome measures, namely a change in diagnosis or diagnostic confidence, utilised in this study have been derived from previously validated measures (24,25,29). Clinicians will be asked to independently select a most likely and most important diagnosis for each case. The most likely diagnosis is self-explanatory; the most important diagnosis is defined as the condition that the clinician would not want to miss in this patient, even if it is very unlikely (29). The diagnoses available for the most likely and most important are identical, and have been compiled based on an audit of most frequent discharge diagnoses in which a CXR was performed. This method of `pruning' has been demonstrated not to adversely bias results (24). Five point Likert scales will be used to measure the clinicians' confidence in their diagnostic decisions, prior to and in conjunction with a CXR report. Each report will be assessed by 2 clinicians of each grade, with post-CXR cases only given to clinicians who had reviewed the initial case summary. This will occur over a minimum of 2 sessions, with 10% duplication to determine intra-observer agreement. The level of intra-observer variation will allow accurate comparison of results between clinicians', accounting for individual characteristics and preferences.

Analysis

Cases which produce changes to a new post-CXR diagnosis or alteration in confidence of an existing decision will be identified. Any report that results in (a) a new diagnosis (most likely/most important) or (b) produces a change in confidence of an existing diagnosis of either 2 points on the Likert scale or a move to an anchor, will be deemed important (24,29,30). The proportion of RR reports producing a positive influence will be compared to CR reports, utilising a pre-defined noninferiority margin of 10%. Analysis of the proportion of RR and CR reports that result in a new diagnosis will be conducted using a McNemar test (36,39). A Sign test will be utilised to determine if there is any difference in the proportion of reports that produce a change in diagnostic confidence between RR and CR reports, using Bonferroni's correction for multiple observers (36), except where already corrected for by cluster analysis. Chi-Squared test will examine if there is any difference in the proportion of RR and CR reports producing a change in diagnosis or diagnostic confidence between clinicians' and clinician grade(36,39). Kappa analysis will examine intra-observer agreement (16,36). Local User Groups will be consulted once the results have been obtained, to discuss interesting trends and overall results. Their views will be incorporated into the analysis, to ensure that the findings and practice recommendations produced include patients.

iii) Reliability & Validity

The data collection methods employed in Phase 1 are modelled on accepted standards (10,13,35). The Phase 2 proformas used to identify the influence that CXR reports have on the clinical reasoning of clinicians have been modelled on previously validated tools and outcome measures (25,28,29). The diagnoses available for the most likely and most important are identical, and have been compiled based on an audit of most frequent discharge diagnoses in which a CXR was performed at the hospital from which the CXR cases were drawn. This method of ‘pruning’ has been demonstrated to not adversely bias results (26). Reliability of all participants in Phase 1 and 2 will be examined through 10% duplication with Kappa used to determine intra-observer agreement (16,37).

iv) Ethical considerations

The anonymised cases used in the assessment of diagnostic accuracy will be obtained retrospectively from images performed for clinical reasons with no additional radiation exposure required. Formal written patient consent is not required.

Trust data protection procedures will be observed at all times. Prior to interpretation the x-rays will be anonymised, and with only gender, age in years, referral source and the clinical history provided at the time of initial examination, being retained. All cases used will be assigned a unique study number for the purposes of this study, and will be kept independent from the study data in a locked filing cabinet within the Radiology offices. Participants will be assigned unique identifiers to ensure anonymity, and will be securely stored separately to the collected data. All participants will be voluntarily recruited and free to withdraw at any time. The physician compiling the case summaries from the patient notes will be an employee of the Trust, produce anonymous information and follow established procedures for handling sensitive patient data at all times.

Any case in which one or both of the expert chest radiologists identify an abnormality not detected at the time of initial investigation will be highlighted by the researcher to the Clinical Directors of Radiology and Respiratory Medicine. The case will be reviewed any subsequent imaging to ensure that appropriate management occurred. If it is decided that a significant abnormality has been missed a Trust incident form will be completed and the patient's treating clinician notified who will contact the patient according to Trust procedure.

Potential impact of study

If RRs are shown to interpret CXR with comparable accuracy to CRs, and there is no clinically significant difference in the influence that these reports have on clinicians' clinical decision-making, RRs could provide an additional reporting resource to the NHS in an efficient and effective way. This would increase the volume and timeliness of reporting, enable streamlined patient pathways and improved patient care, while maximising the limited resources available.

Dissemination strategy

A schedule of articles arising from the proposed programme of research has been developed to disseminate findings to the various stakeholders, and will be guided by discussions with local User Groups, ensuring the outputs are focused on the needs of service users. Agreement between the expert chest CRs and the process of obtaining the robust reference standard diagnosis will form the basis of a paper to be submitted to the `Radiography' journal, as well as an abstract to be proffered to the European Respiratory Society congress. The accuracy with which the RRs and CRs interpret the standardised test bank will contribute to the evidence base for radiographer reporting, and will be submitted to the British Journal of Radiology. The influence that RR CXR reports have on the decision-making of clinicians will address a gap in the current knowledge base, and be of interest to clinicians' that use x-rays as well as those who provide reports, and will be submitted to the British Medical Journal. An abstract outlining the research programme and results will be proffered for presentation at the Radiological Society of North America congress to raise the profile of radiographer reporting internationally. Updates reflecting progress will be submitted annually to the UK Radiological Congress as well as local research meetings in the Trust and University. The robust methods employed in the various arms of this programme can be used as benchmarks for other practitioners interested in performing diagnostic accuracy studies or examining the influence that diagnostic testing has on clinical decisions, both within the radiology-radiographer sector and beyond. Abstracts outlining the research methodologies and data analysis will be offered for presentation at University post-graduate research seminars.

The results of each phase will be presented to the Trust Users and Hackney Local Involvement Network (LINK) Group meetings once preliminary analysis has been conducted. Their views and thoughts will help drive the final analysis direction and the conclusions from each Phase and the study as a whole will be fed back to these patient groups. Relevant findings from each Phase will also be included on open patient information evenings organised by the Trust.

References:

1. Brealey S, Scally A, Hahn S, et al. Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. Clinical radiology 2005;60(2):232-41.

2. Secretary for Health. Equity and excellence: Liberating the NHS. Department of Health UK: Crown Stationary Office, 2010.

3. NHS Improvements. The best of clinical pathway redesign: Practical examples delivering benefits to patients. Department of Health UK, 2011.

4. NHS Improvements. First steps towards quality improvement: A simple guide to impoving services. Department of Health UK, 2011.

5. Department of Health. Radiography Skills Mix: A report on the four-tier service delivery model. Department of Health UK, 2003.

6. Paterson A, Price RC, Thomas A, et al. Reporting by radiographers: a policy and practice guide. Radiography 2004;10(3):205-12.

7. Hollingworth W, Jarvik JG. Technology assessment in radiology: putting the evidence in evidence-based radiology. Radiology 2007;244(1):31-8.

8. Brealey S, Scally AJ. Bias in plain film reading performance studies. BJR 2001;74(880):307-16.

9. Abubakar I, Story A, Lipman M, et al. Diagnostic accuracy of digital chest radiography for pulmonary tuberculosis in a UK urban population. The European respiratory journal 2010;35(3):689-92.

10. Eisen LA, Berger JS, Hegde A, et al. Competency in chest radiography. A comparison of medical students, residents, and fellows. J Gen Intern Med 2006;21(5):460-5.

11. Eng J, Mysko WK, Weller GE, et al. Interpretation of Emergency Department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and film with digital display. AJR. 2000;175(5):1233-8.

12. Herman PG, Gerson DE, Hessel SJ, et al. Disagreements in chest roentgen interpretation. Chest 1975;68(3):278-82.

13. Koppaka R, Bock N. How reliable is chest radiography? In: Organisation WH, editor. Toman's tuberculosis: Case detection, treatment and monitoring. 2nd Edition ed. Geneva: World Health Organisation, 2004:51-60.

14. Potchen EJ, Cooper TG, Sierra AE, et al. Measuring performance in chest radiography. Radiology 2000;217(2):456-9.

15. Quekel LG, Kessels AG, Goei R, et al. Miss rate of lung cancer on the chest radiograph in clinical practice. Chest 1999;115(3):720-4.

16. Robinson P, Wilson D, Coral A, et al. Variation between experienced observers in the interpretation of accident and emergency radiographs. BJR 1999;72:323-30

17. Zellweger J-P, Heinzer R, Vidondo B, et al. Intra-observer and overall agreement in the radiological assessment of tuberculosis. Int J Tuberc Lung Dis 2006;10(10):1123-26.

18. Flehinger BJ, Melamed MR, Heelan RT, et al. Accuracy of chest film screening by technologists in the New York early lung cancer detection program. AJR. 1978;131(4):593-7.

19. Sherman RS. Screening of Chest Roentgenograms by Advanced Roentgen Technologists. Investigative Radiology 1971;6(1):76.

20. Dixon AK. The impact of medical imaging on the physician's diagnostic and therapeutic thinking. European radiology 1998;8(3):488-90.

21. Hollingworth W. Radiology cost and outcomes studies: standard practice and emerging methods. AJR. 2005;185(4):833-9.

22. Dixon AK, Southern JP, Teale A, et al. Magnetic resonance imaging of the head and spine: effective for the clinician or the patient? BMJ 1991;302(6768):79-82.

23. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography. Effect on diagnostic and therapeutic plans. JAMA 1977;238(3):224-7.

24. Hobby JL, Dixon AK, Bearcroft PWP, et al. MR Imaging of the Wrist: Effect on Clinical Diagnosis and Patient Care. Radiology 2001;220(3):589-93.

25. Mackenzie R, Dixon AK, Keene GS, et al. Magnetic resonance imaging of the knee: assessment of effectiveness. Clinical radiology 1996;51(4):245-50.

26. Ng CS, Palmer CR. Assessing Diagnostic Confidence: A Comparative Review of Analytical Methods. Academic Radiology 2008;15(5):584-92.

27. Wittenberg J, Fineberg HV, Black EB, et al. Clinical efficacy of computed body tomography. AJR 1978;131(1):5-14.

28. Wittenberg J, Fineberg HV, Ferrucci JT, et al. Clinical efficacy of computed body tomography, II. AJR. 1980;134(6):1111-20.

29. Lusted LB. A study of the efficacy of diagnostic radiologic procedures. Final report on diagnostic efficacy. Chicago: American College of Radiology, 1977.

30. Brealey SD, King DG, Hahn S, et al. Radiographers and radiologists reporting plain radiograph requests from accident and emergency and general practice. Clinical radiology 2005;60(6):710-7.

31. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326(7379):41-44.

32. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. Annals of internal medicine 2007;147(8):W163-W94.

33. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006;295(10):1152-60.

34. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to cluster randomised trials. BMJ 2004;328(7441):702-8.

35. Obuchowski NA. Sample size tables for receiver operating characteristic studies. AJR. 2000;175(3):603-8.

36. Sardanelli F, Di Leo G. Biostatistics for Radiologists: Planning, Performing, and Writing a Radiologic Study. Italy: Springer-Verlag Italia, 2009.

37. Scally AJ, Brealey S. Confidence intervals and sample size calculations for studies of film-reading performance. Clinical radiology 2003;58(3):238-46.

38. Wears RL. Advanced statistics: Statistical methods for analyzing cluster and cluster-randomised data. Academic Emergency Medicine 2002;9(4):330-41.

39. Hazard-Munro B. Selected nonparametric techniques. In: Hazard-Munro B, editor. Statistical methods for healthare research. 5th ed. USA: Lippincott Williams & Wilkins, 2005:109-35.