The Diagnostic Accuracy of Reporting Radiographer Chest X-ray Interpretations and their Influence on Clinicians' Diagnostic Decision-Making: A Comparison with Consultant Radiologists

by

Nick Woznitza

Canterbury Christ Church University

Thesis submitted for the Degree of Doctor of Philosophy

2016

Abstract

Background: Diagnostic imaging plays an expanding and central role in patients' medical care. Radiographer clinical reporting is being increasingly used in patient focused services. There is a paucity of research that has examined radiographer chest X-ray reporting.

Aim: To determine the diagnostic accuracy of reporting radiographer chest X-ray (CXR) reporting and the influence that CXR reports have on clinicians' diagnostic decision-making.

Method: A quasi-experimental study determined the diagnostic accuracy of a cohort of reporting radiographers in CXR interpretation, using a free-response methodology. The influence of CXR reports on clinicians' diagnostic decision-making was determined with a cohort study. A non-inferiority approach was used, in line with Royal College of Radiologists and College of Radiographers guidance that reporting radiographers must be comparable to consultant radiologists.

Results: The diagnostic accuracy of reporting radiographers (RR) was non-inferior to consultant radiologists (CR) for all measures, all p<0.0001; unweighted JAFROC (RR Figure of Merit [FoM]=0.828, 95%CI 0.808–0.847; CR FoM=0.788, 95%CI 0.766–0.811), weighted JAFROC (RR FoM=0.830, 95%CI 0.811–0.849; CR FoM=0.786, 95%CI 0.764–0.808) and inferred ROC (RR Area Under the Curve [AUC]=0.909, 95%CI 0.887–0.931; CR AUC=0.903, 95%CI 0.882–0.924). No difference was found in the number of CXR reports that produced a correct most likely and/or most serious diagnosis (RR 876 of 1337 cases; CR 810 of 1368; p=0.103). Uncorrected most likely diagnostic confidence (RR 72.5 to 80.2; CR 71.0 to 80.4) and uncorrected most serious diagnostic confidence (RR 34.0 to 41.9; CR 33.5 to 39.2) of reporting radiographer CXR reports was non-inferior to consultant radiologists (p<0.001). Corrected most likely diagnostic confidence, calculated using the Tsushima methodology, was lower (RR 4.61; CR 5.02) with no apparent difference, but non-inferiority was not confirmed (p>0.05).

Conclusion: With appropriate postgraduate education, reporting radiographers are able to interpret chest X-rays at a level comparable to consultant radiologists.

Lay Abstract

Background: Diagnostic imaging plays an expanding and central role in patients' medical care. Radiographer clinical reporting is being increasingly used in patient focused services. There is a paucity of research that has examined radiographer chest X-ray reporting.

Aim: To determine the accuracy of reporting radiographer chest X-ray (CXR) reading and how doctors use different chest X-ray reports in patient care.

Method: One study established the accuracy of chest X-ray interpretation of a group of reporting radiographers. A second study explored the influence that chest X-ray reports had on a group of doctors who use chest X-ray reports when treating patients. The study designs enabled direct comparison with consultant radiologists, to ensure the radiographers were safe practitioners.

Results: The group of reporting radiographers had similar accuracy to the consultant radiologists when reading chest X-rays, and this was statistically significant. The chest X-ray reports of consultant radiologists and reporting radiographers were used in comparable ways by doctors. Some aspects, for example the number of correct diagnoses, were statistically significant. Other measures, for example corrected diagnostic confidence, were broadly equal between the two groups but not statistically significant.

Conclusion: With appropriate postgraduate education, reporting radiographers are able to interpret chest X-rays at a level comparable to consultant radiologists.

Acknowledgements

This research would not have been possible without the unwavering support of my wife, Anne-Marie, and the never-ending patience of my children, Alice and Monty. You have all sacrificed so much for this thesis. Words will never describe my love, or my eternal gratitude.

This research has been part-funded by the College of Radiographers Industry Partnership (CoRIPS) scheme. This support has been crucial in completing the research.

Great thanks is given to the consultant radiologists, reporting radiographers and clinicians who gave their time to participate in the project. Without them, there would be no data to analyse.

My supervisors have nurtured me through this long journey, and for that I am forever grateful; Dr. Keith Piper, Prof. Kate Springett and Prof. Audrey Paterson from Canterbury Christ Church University and Prof. Graham Bothamley and Prof. Narendra Aladangady from Homerton University Hospital. The support given by Dr. Susan Rowe, Dr. Stephen Burke, Mr. Colin West and the radiology team at Homerton University Hospital has been instrumental and has allowed me to develop as a clinician and academic.

And to my parents, Hans and Mary. You both taught me that anything was achievable with hard work and persistence. Looks like you were right.

Table of Contents

Chapter 3 - Methodology

Chapter 4 – Results

Chapter 5 – Discussion

List of Appendices

List of Figures and Tables

Figures

Tables

Chapter 1 – Introduction

If we are ever going to get the 'optimum' results from our national expenditure on the NHS we must finally be able to express the results in the form of the benefit and the cost to the population of a particular activity, and the increased benefit that could be obtained if more money was made available.

A. L. Cochrane (1972) *Effectiveness and Efficiency: random reflections on health services*[1](p.1)

1.1 A historical review of radiographer reporting development and current and future context

Clinical imaging plays a central role in healthcare and is viewed by clinicians as an essential tool to support them in diagnostic and therapeutic decision-making. New and emerging technologies, coupled with a growing and aging population and increasing treatment options, has seen the demand for clinical imaging outstrip capacity within the United Kingdom (UK).[2-4] The situation is not unique to the UK, with service delivery challenges due to sustained increases in imaging echoed worldwide.[5 6] Early and accurate diagnosis, especially for cancer,[7 8] are suggested as methods for improving patient outcome and experience,[9-11] with diagnostic capacity frequently identified as a barrier to achieving these goals.[12] Delayed clinical reports are a serious factor in the diagnostic capacity barrier,[13-15] and training radiographers to undertake clinical reporting has addressed this situation to some extent.[4 16-18] These issues are not a recent phenomenon and, since the National Health Service and Community Care Act (1990),[19] trained reporting radiographers have provided an increasing contribution to clinical reporting.[4 20 21] Initially focused on the reporting of trauma skeletal X-rays,[22] reporting radiographers have expanded their scope of practice to include the complete spectrum of skeletal X-rays,[23 24] mammography,[25-27] gastrointestinal imaging,[28 29] and selected magnetic resonance imaging examinations.[30 31] Radiographer reporting, particularly skeletal reporting, has become widespread across the UK,[4 32] and in many departments provides a significant contribution to reporting capacity.[17 18] Chest X-ray interpretation by radiographers is also not a new concept, with early work conducted as part of the lung cancer screening programmes exploring this in the 1970s.[33 34] More recently, reporting radiographers have been trained to report chest X-rays and are doing so although the practice is not as established as skeletal reporting.[21 35] One reason that chest X-ray reporting by radiographers is not more established would appear to be the relative paucity of research evidence which has examined the diagnostic accuracy of chest X-ray reporting compared to the extensive body for skeletal reporting.[22-24 36-40] Promising results have been reported in two studies which have explored the accuracy of radiographer reporting of chest X-rays, one in an academic setting[41] and the other in clinical practice.[42] No study appears evident in the literature that directly compared the performance of consultant radiologists and reporting radiographers when interpreting chest X-rays, and there appears to be no work which has examined the influence of reporting radiographer chest X-ray reports on clinicians' diagnostic decision-making. Both the comparison of performance and clarity over whether who reports on the chest X-ray influences practitioners' clinical decisions are important, as it is vital that quality and patient safety are maintained if a new service is introduced into practice.[43 44]

The purpose of this study was to compare the diagnostic accuracy of chest X-ray interpretation of consultant radiologists and reporting radiographers, and the influence that chest X-ray reports have on clinicians' diagnostic decision-making.

For the purposes of the current work, the following definition of diagnostic decision-making is used:

The collection of additional information intended to (further) clarify the character and prognosis

(adapted from Knottnerus *et al.* 2009)[45]

This thesis will begin with a discussion of evidence based practice, the hierarchy of efficacy used to evaluate health technologies (which includes radiographer reporting), as well as the existing literature on chest X-ray reporting accuracy and the influence that radiology reports have on clinicians' diagnostic decision-making (Chapter 2). The study consists of two parts. The quasi-experimental diagnostic accuracy methodology for Part 1 has been informed by the STAndards for Reporting Diagnostic accuracy (STARD) framework,[46 47] incorporating the non-inferiority requirements of the CONsolidated Standards Of Reporting Trials (CONSORT).[48] The second part, an observational study that has examined the influence of chest X-ray reports on clinicians' diagnostic decision-making has utilised the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidance.[49] A comprehensive analysis of the design was performed and reported in Chapter 3. The results of the diagnostic accuracy and diagnostic impact studies are presented in Chapter 4, and the findings of the current study are placed into context in Chapter 5.

1.2 Evidence based practice and the hierarchy of efficacy

Adapted from the medical shift to evidence based medicine (EBM) in the 1990s,[50 51] evidence based practice (EBP) has been adopted by nursing[52] and the allied health professions.[53] Evidence based medicine and evidence practice is not without its shortcomings, and recent criticisms have been raised. A notable example is the application of average results from clinical trials with strict inclusion and exclusion criteria to a heterogeneous population of patients encountered in clinical practice with a range of co-morbidities.[54] Another important consideration is the crucial distinction between statistical and clinical significance.[54] Both of these concepts have been considered and addressed within the study design. For radiographer reporting to be evidence based, a robust assessment of diagnostic accuracy is required, to ensure patient safety is maintained if it is introduced into clinical practice. Methodologies for determining the influence of radiology investigations on patient outcomes have been developed, with the hierarchy of efficacy[55-57] adapted into health technology assessment (HTA),[58-60] to provide ordered and structured assessments of new modalities and techniques. This systematic and ordered approach, whereby lower levels of efficacy (technical efficacy, diagnostic accuracy, diagnostic impact) are required to be established to assess higher levels (patient outcome, societal outcome).[57 61] Methodologies and intermediate outcome measures are also proposed, with Brealey[62 63] adapting the existing radiologist centred framework of Fryback and Thornbury,[57] further developed by Mackenzie and Dixon[61] to radiographer reporting (see Chapter 2.6).

1.3 The need for a diagnostic accuracy assessment of radiographer chest X-ray reporting

A substantial body of work has established significant variation in the diagnostic accuracy of chest X-ray interpretation between observers.

For the purposes of the current work, the following definition is used:
Observer: the healthcare professional who interprets an imaging investigation
Adapted from Chakraborty (2013)[64]

Historical analysis of tuberculosis screening programmes during and immediately after World War II reported wide variation between observers,[65-70] with little change in observer performance during the early lung cancer screening trials.[71-73] Despite advances in medical technology, more recent studies still report substantial variation in image interpretation accuracy between observers for tuberculosis,[74-76] lung cancer[77-79] and when reporting chest X-rays in a controlled setting.[80-85]

If radiographers are to provide clinical reports for chest X-rays, this variation in accuracy between observers needs to be considered when comparing the performance of the consultant radiologists and reporting radiographers. The current study has investigated the diagnostic accuracy of chest X-ray reporting by reporting radiographers, and the influence that these reports have on clinicians' diagnostic decision-making. Direct comparison with consultant radiologists has been made, which has allowed a decision regarding suitable performance of reporting radiographers to be made. The results of the diagnostic accuracy study (Chapter 4.2 and Chapter 4.3) and the diagnostic impact study (Chapter 4.4) have been placed in the context of the existing literature (diagnostic accuracy Chapter 5.2 – 5.4; diagnostic impact Chapter 5.5) and have determined that trained reporting radiographers interpret chest X-rays with comparable accuracy to consultant radiologists, with no detrimental influence on clinicians' diagnostic decision-making.

1.4 Need for radiographer reporting – diagnostic capacity

Imaging demand has outstripped diagnostic capacity worldwide. With new and emerging techniques,[86 87] coupled with novel uses of existing technologies,[88] radiology has become embedded into an increasing range of patient pathways. Different health environments have different challenges, from limited equipment capacity,[89] minimal reductions in waiting times despite capital investment[5] and shortages of trained healthcare professionals to interpret the investigations.[4 6 14 15 90] Exacerbating the situation is that imaging examinations are becoming quicker to acquire as technology advances, and are becoming more complex,[91] with an increasing focus on interventional radiology as the preferred treatment option.[92] For health economies without access or reporting constraints, cost pressures are becoming a concern.[93] Service challenges on radiology in the United Kingdom are threefold; sustained increases in activity,[9 94] a chronic shortage of consultant radiologists[4 95] and unprecedented economic restrictions.[96]

It is in this climate that radical changes to the provision of healthcare[11] and cancer pathways[8 97] in the UK are being implemented. Recognising the need to improve patient outcomes for cancer, especially lung cancer which has shown stagnant survival rates,[98 99] renewed focus is being given to rapid referral and diagnosis in cases of suspected cancer.[8 100 101] These initiatives will undoubtedly increase the volume of imaging investigations performed, at a time when diagnostic capacity is failing to meet current demand.[12] It is clear that the status quo for radiology service delivery is no longer an option, and it is a fundamental requirement that all healthcare professionals within the diagnostic pathway operate at their maximum ability and skill set to meet current and future demand in a patient focused way.[102 103] There is some evidence that radiographer advanced practice, which incorporates radiographer reporting of chest X-rays, can help to meet these challenges,[17] but this has yet to be adopted on a wider scale which may be limiting benefits to patients.[32 35]

Chapter 2 – Literature Review

2.1 The history and evolution of evidence-based medicine

Gordon Guyatt is often credited with the first use of the phrase "Evidence-based medicine" in his 1991 article of the same name,[50] although one of the most widely encountered definitions in the literature is that of Sackett *et al.* who states that evidence-based medicine (EBM) is "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients"(p.71).[51] Several other authors have developed their own interpretation of what in their views constitutes evidence-based medicine. All place the use of high quality research evidence at the heart of patient care.[63 104-110] It is clear from these definitions that the primary aim of evidence- based medicine is to focus the clinician to emphasise the role that research evidence should have on everyday practice. Systems to collate the vast quantities of evidence in a systematic way have been developed. The GRADE methodology (Grading of Recommendations, Assessment, Development and Evaluation) which has become one of the most widespread reviews the quality of evidence for a specific clinical outcome and aligns recommendations with the strength of the supporting evidence.[111 112]

Since its origins in the early 1990s, evidence-based medicine has quickly become the accepted method of practising medicine, although it has not been universally accepted or implemented without challenge.[54] The concept of evidence-based care has expanded from its initial medical focus to encompass the entire healthcare system. Embraced and adapted by many allied health professions,[53 104] evidence-based medicine has evolved to become the concept of evidence-based practice in recognition of its impact on the way best care is delivered. This is evident in the explicit emphasis placed on the principles of evidence-based medicine and evidence-based practice as related to the standards of proficiency for allied health professionals within the United Kingdom (UK), including radiographers.[113]

Evidence-based medicine is not a new concept. Although the term appeared, as we understand it, in the early 1990s, this was just seen as formalising and defining a system of practice that had been evolving for nearly 20 years.[104 114-116] Some argue that medicine has always been evidence-based, but the proponents of EBM have adjusted the focus to the quality & reliability of the evidence that is used when making diagnostic and therapeutic decisions.[109 116 117] The evidence-based medicine movement arose from the need to change the type of evidence used and the way this evidence was integrated into care. The seminal paper often cited as the birth of clinical epidemiology was the 1836 study on the effectiveness of bloodletting in pneumonitis.[118] Louis based his assertions on the efficacy of therapy and patient outcome from conclusions drawn from the systematic observation of multiple cases.[118] This was one of the first examples of an objective analysis of medicine and the effectiveness of therapy, which caused controversy by challenging the established traditions of the time.[117 119] It is this emphasis on objective rather than subjective evidence that clearly defines the practice of evidence-based medicine.

The deeper historical roots allowed the structure of EBM to grow during the 1970s, spurred on by the explosion of research information available. This required practising clinicians to develop the skills required to effectively search and digest primary research data in order to provide optimal care.[107 115] Effective procedures for literature searching were developed and validated by Haynes,[120] and the key concepts adapted to the clinical environment by Guyatt.[116] These were then formulated into the model of EBM; solving a specific and answerable clinical question through a critical evaluation of a systematic literature search.[63 121-123] Readily identifying biases and weaknesses in study design allows clinicians to rapidly and accurately assess the quality of research by highlighting the key methodological concepts that underpin strong evidence, and several authors have developed and distributed tools allowing practitioners to critique published research.[46 63 105 121-123]

There is little doubt that evidence-based practice is deeply entwined within an empiricist perspective of reality. Evidence-based practice values the results of robust, controlled clinical trials over the subjective observations of practitioners, uncontrolled case studies or inferences taken from pathophysiological models.[51 107-110 116 119] There are many who would question this fundamental aspect of evidence-based care. As the selection of objective measures is of itself, a purely subjective task, from a philosophical and epistemological standpoint, is there any such thing as truly objective knowledge?[117 124] Other authors, such as Ashcroft and ter Meuken for example,[114] would take this one step further and pose the question, does evidence equate to knowledge? Does the evidence-based revolution define a new nature of knowledge? It is within these deeper questions that many critics of evidence-based practice find issue.[109]

Evidence-based medicine and evidence-based practice favours scientific results over expert recommendations.[51 111 125 126] The delineation between scientific and non-scientific evidence is made on whether a hypothesis has been tested.[127] A hypothesis is a statement that is proposed, often an association or effect, which is tested within an experiment by controlling a set of variables.[127] The results would then be used to infer that the hypothesis should be accepted or rejected. Central to this was the work of Popper and the notion of falsifiability.[127] In essence, the concept is centred on the notion that it is difficult to prove something is right, but much easier to prove something is wrong. Philosophically this process has evolved to acceptance or rejection of a null hypothesis. As discussed in Appendix 12.6, the notion of non-inferiority draws on the concept of falsifiability and produces a reversal of the traditional alternate and null hypothesis structure common to medicine and science.

The adoption of evidence-based medicine has been described as a paradigm shift.[125] A paradigm, according to the philosopher Thomas Kuhn, is an entire system of beliefs, practices and values shared by a community of scientists that is used to explore and explain reality.[128] Acceptance of core fundamentals within a paradigm allows scientists to concentrate their efforts at the periphery, attempting to clarify irregularities within theories. It is this process which is described as 'normal science'(p.77).[124] If these minor inconsistencies accumulate or major deficiencies are identified within a paradigm. Major deficiencies can produce a crisis, the catalyst for a scientific revolution. Revolutions result in a fundamental shift, the acceptance of a new paradigm, a new set of assumptions and a new 'world view'(p.16).[109] By definition this new paradigm is at least in part, if not completely, incommensurable with the old paradigm.[124 128]

A vague definition of evidence-based medicine and the implications associated with the term 'paradigm shift' are highlighted by critics as substantial flaws in the fundamental philosophy of evidence-based practice. While at first glance it would be difficult to argue against a system of practice that values best evidence, there is heated debate as to what constitutes best evidence and evidence in general.[109 114 116 117] Choosing to include the term evidence-based within the title raises tensions within the healthcare community. By designating one system of practice to be based on evidence it, by definition, implies that all alternative models are based on something that is not evidence.[117] Highlighted in their critique, Sehon and Stanley postulate that the champions of evidence-based practice have themselves produced no new theory of evidence.[109] What is debated is not the presence or absence of evidence within a paradigm, but the emphasis that should be placed on different types of evidence being utilised.

Originally conceived as a methodology to allow clinicians to digest, interpret and apply the rapidly expanding body of primary research, evidence-based practice (EBP) requires a thorough knowledge of research methodologies and a substantial time commitment.[108 110] Recognising the limitations which can occur in clinical practice, such as workload pressures and system constraints including resource availability and funding, evidence-based care has further evolved to incorporate a top-down element comprising of systematic reviews and evidence based guidelines, often produced by governments and academic centres.[108 110] Critics highlight this fundamental shift in the philosophy and methodology of EBP as reactionary and not well managed within its existing frameworks.[109 116] The core tenets of individual practitioners answering scientific questions through a systematic literature search of primary studies initially seem at odds with governmental led clinical guidelines. In response, the GRADE system was developed as a way to streamline the process of scientific review for defined clinical questions and to reduce some of the burden on individual practitioners. Some perceive the production of guidelines as an attempt to ration services and cut costs, highlighted by many as a weakness within evidence-based practice.[51 107 115] This is often portrayed as an attempt to limit services. These conclusions are robustly refuted by EBM advocates who declare care is careless and a waste of finite resources when it has not been proven to be efficacious and effective.[51 108]

Concerns are also raised that as well as restricting cost, evidence-based practice attempts to limit the clinical freedom and professional autonomy of clinicians, resulting in a proscriptive approach to care. Sehon and Stanley paint the picture of the art vs. the science of medicine, emphasising the importance of experience and clinical judgement,[109] a claim supported by Aveyard and Sharp.[104] These feelings are shared by many critics of EBP, arguing that the excessive focus on research results limits the invaluable contribution that professional judgement must play.[110 116]

An important consideration when interpreting research findings is the distinction between a statistically significant result and a clinically important difference. In their discussion, Mellor and Knapp stress that statistical significance is more aligned with the sample size of a study and the statistical methods used.[129] A result may show significance according to statistical conventions but taken out of context may be practically meaningless. A situation in clinical imaging for example, where a new contrast medium demonstrates a statistically significant increase in attenuation within a tumour by a single Hounsfield unit on computed tomography (CT) would not have any impact on diagnosis. Conversely, if a contrast media achieved the same attenuation but with lower renal toxicity then this would improve care. Sehon and Stanley argue that it is not possible to test hypotheses in isolation, and that statistical results from a clinical trial must be able to be integrated into a basic understanding of physiology and pathology in order to give clinical meaning to the findings.[109]

While it is essential that practitioners be aware of the latest research findings, it is of vital importance that clinical experience and professional judgement be used when assessing the applicability of research findings to individual patient care. While a novel treatment may sound promising, clinical trials often impose strict inclusion criteria that may severely limit the generalisability of the conclusions. It is important that the responsible practitioner integrates these findings and adapts them to the unique situation presented by each individual patient.[51 104 105 107 108 117] It is no longer within an individual practitioner's discretion to apply the principles of evidence-based practice when delivering care as this is now a mandatory requirement set out by the regulating and professional bodies within the United Kingdom.[52 53 130]

Another contentious issue when debating evidence-based practice is the role that the patient assumes within the framework. Patients, patient choice and patient values should be central to any system of care, and this position is taken up simultaneously by the supporters and opponents of evidence-based care. Wyer and Silva draw attention to the seeming lack of patient focus in the various definitions given of evidence-based medicine, a position that is staunchly denied,[116] with Sackett *et al.* firmly placing the emphasis on the individual patient.[51] A patient-centric approach is reinforced by the Health and Care Professions Council, who state that optimal patient focused care should be delivered by using an evidence-based approach.[53]

2.2 Evidence-based radiology and randomised controlled trials

I consider much less thinking has gone into the theory underlying diagnosis, or possibly one should say less energy has gone into constructing the correct model of diagnostic procedures, than into therapy or prevention where the concept of "altering the natural history of the disease" has been generally accepted and a theory has been evolved for testing hypotheses concerning this.

> A. L. Cochrane (1972) *Effectiveness and Efficiency: random reflections on health services*[1](p.35)

Areas which are not handled well by the mantra of evidence-based practice are the facets of care for which there is incomplete or poor evidence. The advocates of EBP are clear in what they see as best evidence, with the randomised controlled trial (RCT) held at the top of the ladder in terms of robust research methods. Controlling for sample selection and potential clinician bias, the randomised controlled trial is one of the few ways of providing robust evidence of improved patient outcomes related to an intervention. The results of clinical trials, meta-analyses and systematic reviews can be applied to many aspects of practice but there are still significant areas in which there is a paucity of this robust evidence.

It has been suggested that a randomised controlled trial is the only way to provide direct evidence of an improvement in patient outcome.[58 131 132] Fryback and Thornbury highlighted that "it is difficult, if not impossible, outside of a prospective randomized controlled trial to attribute some portion of improved patient condition to the use of an imaging test" (p.91).[57] While an RCT produces robust evidence, it is not the appropriate method for investigating all aspects of healthcare. Several authors have commented that due to the very nature of diagnosis, screening and prevention the questions posed in these fields can rarely be answered by a practical randomised controlled trial.[107 108] It has been hard to reconcile the key principles of evidence-based practice into a diagnostic environment such as clinical imaging.

Although highly valued, randomised controlled trials of diagnostic investigations are often problematic and not well suited to everyday practice. Arising from a therapeutic standpoint where nearly all of the participants have the target condition the sample sizes required to obtain appropriate power can rarely be obtained in clinical imaging. The spectrum of disease for which many radiology procedures are utilised, as well as the large temporal distance between initial investigation and ultimate patient outcome all pose significant obstacles when conducting trials within clinical imaging. Another confounding factor is the ethical and practical inability to effectively 'blind' the treating clinicians to the results of an investigation making the creation of an effective control group difficult.[57 61 63 131 133 134] For diagnostic tests that have already become part of routine clinical practice, patient and clinician resistance and ethical concerns of withholding a test also need to be considered.[135 136]

Some examples of where well conducted, adequately powered randomised controlled trials have been performed to investigate radiology investigations are screening mammography[137 138] and lung cancer screening.[139 140] Evaluations of screening for disease lend themselves to randomisation as the population, by definition, are asymptomatic which removes some of the ethical boundaries to withholding investigation.[135 136] Unlike routine radiology investigations that are used to investigate a broad range of pathologies, detecting disease with an effective treatment in asymptomatic patients also removes some of the confounding factors from the study design. This allows benefits on patient outcome (mortality, morbidity) to be attributed directly to the imaging intervention.[137 139]

Randomised controlled trials are even less frequent in radiographer reporting. A recent study has however established the positive impact on patient outcome and the cost-effectiveness of a radiographer-led emergency department immediate skeletal reporting service, with few radiographer errors and a reduction in both patient recall, unnecessary treatments and total costs.[141] The targeted nature of the intervention, namely immediate radiographer reporting of skeletal X-ray examinations in an emergency care setting, and the proximity of the intervention to the outcome, help to achieve adequate recruitment and sufficient sample size. For radiographer reporting in other areas that have a potential impact across a wider spectrum of diseases, this may pose a significant barrier. For example, chest X-rays have a role in the diagnosis and management of a range of diseases and patient pathways, such as cancer, chronic obstructive pulmonary disease and infection, and it is more difficult to select a population or disease a priori when designing a study.

It is for these reasons that evidence-based radiology (EBR) has been relatively slow to evolve, with the first guideline publication from the evidence-based radiology working group arriving in 2001.[108] In recognition of the gaps and limitations in available research evidence within radiology, yet with the requirement for evidence-based practice, the contribution that well conducted observational research provides has been recognised. In the absence of RCTs, observational studies provide a method of examining the efficacy, effectiveness and efficiency of radiology.[57 131 134]

2.3 Hierarchy of efficacy and patient outcomes

For many years, much of the research arising from the medical specialties has tended to focus on local goals. The vast majority of radiology research has concerned technical developments and diagnostic accuracy with limited publications examining the impact that imaging has on patient health.[57 59 134 142] The advent and rise to prominence of evidence-based practice has resulted in renewed emphasis being placed on establishing the influence that radiology has on patient care and the role that imaging plays within the patient pathway,[58 59] often described in the literature as diagnostic efficacy.

<u>For the purposes of this study, the following definitions are used</u>:

Efficacy: the influence that a diagnostic investigation has within an ideal (controlled) environment

Effectiveness: the influence that a diagnostic investigation has within normal clinical practice

(adapted from Wittenberg *et al.,*[143] Fryback and Thornbury,[57] Mackenzie and Dixon,[61] Brealey[62] and Sardanelli and Di Leo.[144])

Like evidence-based practice, the notion of diagnostic efficacy is not a recent phenomenon. The early foundations of diagnostic efficacy can be traced to the work examining the impact of radiology on a clinician's decision-making, with Lusted an influential figure.[57 145-148] Initially paralleling the early development of evidence-based medicine, diagnostic efficacy also rose to prominence during the 1970s. Although they started as unique concepts, both share similar fundamental principles and as they developed, they have become inextricably linked.[59] Rapid growth acted as the catalyst for both to develop, evidence-based medicine due to the vast body of primary research available to clinicians,[104 116] and efficacy due to the development of novel, high cost imaging and the corresponding escalation in the cost of healthcare.[56 57 131 149 150] Central to both frameworks is the need for high quality evidence to establish the impact of practice on patient health.[58]

Although the American College of Radiology had commissioned significant work establishing the efficacy of many routine X-ray procedures,[145] the key driver within radiology was the development of computed tomography (CT). This exciting new technology promised to revolutionise the way that diagnoses were made and treatments planned, but this was not without significant expense.[149] If this new technology was to be funded, robust research evidence was required that demonstrated that CT was capable of producing an improvement in patient outcome from this new modality.[131 134]

Radiology investigations generate diagnostic information for clinicians, but this is only one aspect influencing outcome. The biggest influence on overall patient outcome is the availability of effective treatment, with other confounding factors including patient preferences and compliance and the results of other investigations.[59 61 129 133 134 142 151-155] Other considerations unique to diagnosis not encountered in assessment of therapeutic efficacy are the relatively low disease prevalence often encountered, together with the spectrum of disease that can be identified by a given imaging technique.[57 131 134 156]

To overcome these barriers a hierarchy of efficacy was constructed, forming a chain that allowed the impact of radiology at different points within the patient pathway to be assessed. Formalised by Fineberg *et al.*[56] when examining the influence of cranial CT scans on patient diagnosis and management, this draws upon the chain of events proposed by Donabedian when assessing quality in healthcare, distinguishing between structure (resources available), process (the activity of care) and outcome (benefit to patient).[55] The vital and most fundamental aspect of this hierarchy is the reliance on the lower levels of efficacy being established before any benefit at higher levels can be determined. In order for an investigation to be effective, it must first be shown to be efficacious, but the inverse is not true. A new radiology investigation can produce a more accurate diagnosis, but this does not guarantee that there will be a positive influence on overall patient health.[56 57 61-63 134 136 142 157] Technical output, diagnostic information, therapeutic impact and patient outcome comprised the original four tiered structure of Fineberg *et al*. [56] (Figure 2.1) which was widely accepted, embraced, disseminated and adapted by several authors.

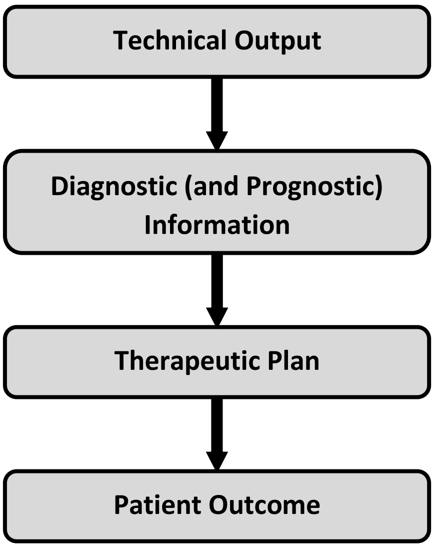


Figure 2.1 The four tier hierarchy of efficacy outlined by Fineberg *et al.* (1977)[56]

Continuation of developments and refinements culminated in the seminal work of Fryback and Thornbury who first proposed the six level hierarchical structure that dominates today (Figure 2.2).[57] Separating the ability of trained observers to reach the correct diagnosis (diagnostic accuracy) from the impact that the radiology report had on a clinicians' diagnostic decision-making (diagnostic impact) was one of the key milestones in this work. Another important contribution was the addition of the highest level, societal efficacy, to reflect the influence that radiology has beyond the individual patient and the need to ensure that limited health resources are used efficiently.

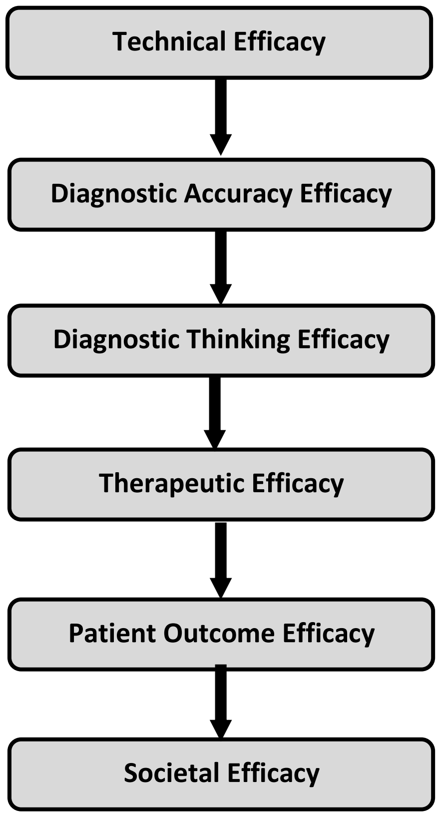


Figure 2.2 The six tier hierarchy of efficacy proposed by Fryback and Thornbury (1991)[57]

Patient outcome is often measured by high-level generic concepts such as morbidity, mortality and more recently quality of life.[58 129 134 149 158-161] While the influence of imaging on eventual patient outcome may not be immediately obvious, examining the sequence of events highlights the valuable role imaging plays. In a typical pathway, the patient is examined by the clinician who determines the need for further investigation and requests the radiological examination. Once the images have been acquired, a trained observer offers their interpretation and produces a report for the clinician, who combines this with the clinical findings and the results of other tests to formulate a diagnosis and create a management plan.[162-164] The principal influence of radiology is the interpretation of the images and the impact of the report on clinicians' diagnostic decision-making, and measuring this effect requires analysis at a transitional outcome level.

In order to evaluate the component levels of efficacy, intermediate outcome measures are required. Intermediate outcome measures are proxy measures of patient outcome and are critical in the evaluation of radiology. By separating out the role that diagnosis and treatment play in the patient pathway, and having the ability to measure the impact of a diagnostic investigation adjacent to the intervention, intermediate outcome measures limit the potential for confounding.[57 59 61 62 129 131 133 136 142 150 153 156] Using intermediate measures in outcomes research utilises deductive logic; in order for patient outcome to be influenced by an investigation it must first change the clinician's diagnostic thinking enough to initiate a change in planned management.[145-147 152 163 165]

The concept of management change as a justification for an imaging investigation is not alien to radiology, with a common question asked by radiology on receipt of an imaging request being 'and how will this impact your management of this patient?'[149]. According to the Ionising Radiation (Medical Exposure) Regulations (IR(ME)R 2000),[166] it is a legal requirement that all requests for medical exposures be justified by a practitioner. The Ionising Radiation (Medical Exposure) Regulations[167] and notes on good practice published by the Royal College of Radiologists (2000), give advice and criteria for medical imaging investigations.[166] When considering the appropriateness of an examination, the practitioner must consider the efficacy of the procedure for the given medical situation,[166] and has been reinforced in a recent review of responsibilities under IR(ME)R (2000).[168]

Each tier on the hierarchy of efficacy is associated with proxy measures of patient outcome, with a list of common measures shown in Table 2.1.

Hierarchical Tier		Common Intermediate Outcome Measures
Level 1	Technical Capability	Resolution, Signal-to-Noise ratio
Level 2	Diagnostic Accuracy Diagnostic Performance	Sensitivity, Specificity, Accuracy, Receiver Operator Characteristic (ROC) curves
Level 3	Diagnostic Impact	Likelihood ratios, Displacement of other investigations, Diagnostic confidence
Level 4	Therapeutic Impact	Change in management plans
Level 5	Patient Outcome	Morbidity, Mortality, Quality of Life (QoL), Quality Adjusted Life Years (QALY)
Level 6	Societal Impact	Economic assessment

Table 2.1 Common intermediate outcome measures adapted from Fryback and Thornbury (1991)[57] and Mackenzie and Dixon (1995)[61]

Although intermediate measures have become accepted when examining the clinical impact of radiology, the use of proxy measures needs to be acknowledged within study designs. As with any proxy measure, diagnostic intermediate outcome measures require validation and, if incorrectly selected, can have limited external validity.[169] Traditionally, the bulk of radiology research has focused on the local goal of radiology, producing anatomically representative images that enable trained observers to make an accurate diagnosis, the lower levels of the hierarchy.[59 61 108 131 142 170] There are relatively few studies that examine the intermediate levels within the hierarchy. Much of the influential work has been conducted on the investigations that have relatively high expense, either in radiation dose, such as CT,[56 143 171 172] or cost, such as magnetic resonance imaging (MRI).[158 162-164] These assessments of efficacy also tended to be conducted shortly after the new technology was introduced into clinical practice, when the volume of requests were small enough to allow prospective studies involving consecutive patients to be conducted. A notable exception is the recent work of Ng *et al.* who examined the impact of early computed tomography of the abdomen and pelvis on the diagnostic work up and management of acute surgical patients, 30 years after the introduction of CT.[172] The aim of Ng's *et al*. study was to determine if the timing of CT for patients with acute abdominal pain influenced the length of hospital stay and accuracy of diagnosis. Early CT was found to reduce not only the length of stay in hospital for this cohort of patients, but also reduced mortality. Given the increasing pressure on health resource, rapid and accurate diagnoses obtained from abdomen and pelvis CT can facilitate prompt management for patients requiring surgical intervention and allow safe and early discharge for those patients that do not require surgery.

2.4 Diagnostic decision-making

Aveyard and Sharp use the definition of decision-making taken from Standing (2005) as "a complex process involving information processing, critical thinking, evaluating evidence, applying relevant knowledge, problem solving skills, reflection and clinical judgement to select the best course of action which optimises a patients'/clients' health and minimises any potential harm..." (p.34).[104] Decisions, whether they are clinical or radiological, are choices made in situations of uncertainty.[145 152 156] In order to reduce uncertainty, clinicians' seek the additional information provided by radiology, with the aim of reducing doubt to a level that optimises therapeutic choices. Kassirer notes however that "absolute certainty in diagnosis is unattainable, no matter how much information we gather, how many observations we make, or how many tests we perform"(p.1489).[152] It is recognised that many clinicians' are uncomfortable with the concept of uncertainty, and often behaving as though it does not exist. It has been suggested that the drive for diagnostic certainty can lead to excessive and redundant investigations.[152 163]

Blanchard *et al.* comment that it is not immediately obvious how a radiological investigation influences patient outcome.[162] The chain of events that surrounds imaging begins with the referring clinician selecting a patient for an imaging investigation. Images are acquired, generating information. The information is then interpreted by trained observers, whose findings are communicated back to the referring clinician, typically in the form of a written report. Subsequently this information is integrated with the clinicians' clinical impressions and the results of other investigations to formulate a diagnosis and determine a management plan for the patient.[61 162 163] It is at the level of clinicians' diagnostic decision-making that radiology, and more specifically the radiology report, has the most measurable impact on patient outcome.[57] Several authors conclude that in order for an examination to influence patient outcome, it must first convince the clinician to do something different – be that initiate, change or withhold treatment.[57 149 150 156 163] The characteristics of clinicians, the relative risk-benefit profile of potential management and patient preferences all influence this decision-making task, with different levels required for different decisions.[104]

2.5 Diagnostic thinking efficacy: the influence that radiology reports have on clinicians' diagnostic decision-making

Early work conducted by the American College of Radiology (ACR) efficacy committee recognised the difficulty in determining the contribution that radiology investigations had on patient outcome.[145] The efficacy committee, whose remit was to determine the efficacy of radiology investigations in medical diagnosis defined three goals; to refine the definition of efficacy, to develop a method(s) of measuring efficacy and to examine efficacy of radiology in medicine.[145] The need for a randomised control trial with long term follow up to determine the influence that intravenous urograms had on patient outcome, with all of the complexities and difficulties associated with the study design and conduct, galvanised the requirements for a step-wise change in the method of assessment.[145]

The largest study that has examined the effectiveness of plain imaging is the 1978 work of the ACR, which examined the diagnostic impact of the most frequently performed plain imaging procedures within 48 accident and emergency (A&E) radiology departments across the United States.[145] Originally intending to investigate the impact that radiology made on patient outcome, the difficulties encountered in several pilot studies led the authors to examine the effect that radiology had on diagnostic efficacy, the level closest to the investigation. Although the ACR determined that plain imaging does have a positive impact on clinicians' diagnostic confidence and decision-making, the most important aspect of the study was that it established that the influence of radiology on clinicians' decision-making could be measured. Such analysis requires the clinicians' pre and post investigation differential diagnoses to be indicated and the confidence in each recorded, enabling likelihood ratios to be calculated. The use of likelihood ratios is an adaptation of Bayes theorem of probability – the influence of additional diagnostic information when making decisions in uncertain conditions.[133 136 145-147]

Assessing clinicians' pre and post investigation diagnoses and confidence as an intermediate outcome measure has been used as the basis for evaluating the efficacy of radiological investigations. Within the current literature, this has been termed diagnostic impact.[61 62] Diagnostic impact is the layer within the hierarchy that links imaging to patient management, acknowledging the role that the clinician plays within the patient care pathway, and is the first step that incorporates influences external to radiology.[57 136] It is crucial that an investigation positively influences diagnostic decision-making without having a detrimental effect on patient outcome.[58] This important consideration has shaped the study design and methodology (Chapter 3.3), with the results presented in Chapter 4.4.

Freedman recognises the highly subjective nature of clinicians' diagnostic assessments,[131] a "value laden exercise" according to Fryback and Thornbury (p.91),[57] although they also suggest that determining a change in clinicians' diagnostic thinking within clinical practice is an empirical question. Hobby *et al.* acknowledge the tendency of observational studies of diagnostic impact to overestimate the benefit of radiology, and recommend using only those diagnoses which are changed to either extreme of confidence (definite or very unlikely) in an attempt to mitigate this.[163] Failure to capture the reassurance to clinician and patient if a serious diagnosis is excluded, or the impact of a normal result in the workup of a patient are both difficult to establish when using standard measures of diagnostic impact.[57 147 152] Another shortcoming of utilising a change in, or confidence in, a diagnosis as a measure of efficacy is raised by Mackenzie and Dixon who acknowledge that "Diagnosis is not an end in itself: only a mental resting place for prognostic considerations and therapeutic decisions" (p.515).[61]

Therapeutic impact, the adjacent level within the hierarchy, captures the influence the radiology investigation and report has on the subsequent management plan of the patient, and is intimately related with diagnostic impact.[56 57 61 62 131 142 149 173] It is at this level where the influence of radiology is heavily mitigated by external factors that will ultimately determine the net benefit of an investigation to overall patient health. While a radiology report can be very accurate and convincing to a clinician regarding the presence of disease, any improvement in patient outcome is limited by the availability of an effective treatment of the condition identified.[56] There are, however, certain circumstances when the new technology demonstrates such clear potential and patient benefit that it becomes unnecessary to examine the higher levels on the efficacy hierarchy. One such example of this rare situation is the impact that cranial computed tomography has in the management of traumatic head injuries, where a highly accurate diagnosis (intracranial haemorrhage) is paired with an effective treatment (neurosurgery).[59]

Outside of a large, prospective randomised controlled trial, it is often very difficult to allocate any improvement in patient outcome to an imaging investigation.[57] There are, however, several well-known limitations when designing and conducting such studies in radiology. It is very difficult, and potentially unethical, to effectively blind a treating clinician to the results of an imaging investigation,[149 174] and the timescale and resources involved are often prohibitively high.[134 159] Even the most ardent advocates of randomised controlled trials recognise these limitations, and acknowledge that they are not a suitable method for examining all aspects of medicine.[51 153] As a result, observational study designs represent the most frequent method for examining the therapeutic impact of radiology investigations.[133 175] The dominant method of measuring therapeutic impact is comparing the management plan of the clinicians when requesting the imaging investigation with the actual treatment delivered through analysis of patients' notes after an appropriate follow up period.[56 143 158 163 171 172 176] Alternative methods for examining therapeutic impact have been proposed, including comparing radiotherapy treatment plans produced with and without access to CT information[177] and by direct questioning of clinicians as to the influence that the investigation has on their management plans.[131] Common criticisms of other methods include the unrealistic nature of planning treatment without access to an accepted imaging procedure and the questionable validity of measuring the hypothetical impact that one facet of the diagnostic pathway has on the eventual treatment of the patient.[57 131 133 145] Determining patient treatment by direct examination of the notes aims to overcome these limitations, and lead to valid and reliable results. Similar to the circumstances encountered when assessing the impact of the normal report on diagnostic impact, it can be argued that a report can produce no change in expected treatment (low therapeutic impact) yet still have a positive impact on patient outcome if the results of the investigation confirm optimal management.[57]

Hillman acknowledges the many potential limitations when considering outcomes research, commenting that "not everything can be studied"(p.s75),[151] while Thornbury suggests that in depth, timely and costly analyses should be reserved for the high cost modalities such as CT and MRI.[159] It could be argued that excluding frequent, high volume investigations from such rigorous analysis is counter-intuitive as the accumulative impact of these investigations is quite high, a feature acknowledged in the analysis by Loop and Lusted.[145] Timeliness is an important factor to consider when planning and conducting these studies, especially in a field that is subject to rapid changes in technology such as radiology.[108 151] Hunink and Krestin have developed a pragmatic methodology that simultaneously examines multiple levels within the hierarchy; the distinct advantage of this approach is the speed with which such appraisals can be conducted while recognising the need for robust evaluations.[58] Another common pitfall when utilising the intermediate outcome measure of diagnostic impact is that the estimates of pre and post investigation diagnosis and diagnostic confidence are subjective measures of clinicians' attitudes rather than objective measures, although as mentioned earlier Fryback and Thornbury suggest that determining a clinician's change in diagnostic thinking in clinical practice is an empirical task.[57] The limitations of subjective measures are well documented, with Blackmore *et al.* commenting that "physicians do not always behave as they say they will" (p.s13),[133] a problem that can be overcome by direct examination of the patients' records taken from clinical practice.

The concept of diagnostic efficacy has been embraced within evidence-based practice, becoming one of the key tenets through its incorporation in health technology assessments (HTA).[59 61 134 150] Brealey has defined healthcare technology as "methods used by healthcare professionals to promote health, prevent and treat disease" (p.341).[62] Technology assessments have been described as the backbone of evidence-based radiology, a fundamental requirement when assessing diagnosis according to evidence-based medicine principles.[57 108] This is further emphasised by the observation of Hollingworth and Jarvik that evidence-based radiology is only as good as the technology assessment on which it is based.[59] Originally the primary focus of technology assessments was to produce high quality data for use by clinicians'. More recently within the United Kingdom the emphasis has shifted to providing evidence for use in guidance and guideline production which will be utilised by government when considering central funding.[59 60] An important consideration when assessing the strength of research relating to radiology is the different levels required by different consumers, with individual practitioners requiring a different level of evidence regarding efficacy when compared to Government considering funding a new or existing technology, a fact acknowledged within the HTA framework. As Hollingworth and Jarvik note, it is easier for government to withhold funding for a new technology that has yet to be implemented than it is to withdraw it once a formal assessment has taken place.[59]

Technology assessments provide a strategy for researchers to ensure primary research studies produce robust evidence, thus reducing the amount of care based on inadequate evidence.[59] Within the United Kingdom, the National Institute for Health Research (NIHR) Health Technology Assessment programme outlines the fundamental questions that a technology assessment is required to answer:

1. whether the technology works
2. for whom
3. at what cost
4. how it compares with the alternatives

   adapted from National Institute for Health Research 2012[178]

The National Institute for Health and Clinical Excellence (NICE), responsible for the production of guidance on which governmental funding is based within England and Wales, is explicit in its requirements of technology assessments. All technology assessments must provide robust evidence that a new technology provides “equivalent or enhanced clinical outcomes for equivalent or reduced cost” (p.10) if it is to be considered for use in the National Health Service (NHS).[60]

2.6 <u>Adaptation of the hierarchy of efficacy to the United Kingdom and radiographer reporting</u>

Treating radiographer reporting as a health technology means that the principles of the hierarchy of efficacy can be applied to it. Adapted concepts at each level of the hierarchy would be required and, in any evaluation of the efficacy of radiographer reporting, a series of questions would be posed and examined.[62 63] Adaptation from the traditional radiologist-centred model was required because, as the established profession for interpreting medical imaging, radiologists are not generally required to prove their accuracy in comparison to another professional group. Radiographer reporting, although established for more than 20 years, is still a change from the traditional model of service delivery. As such, reporting radiographers are required to establish that their performance is comparable to consultant radiologists so that patient safety and outcome is not compromised.[44]

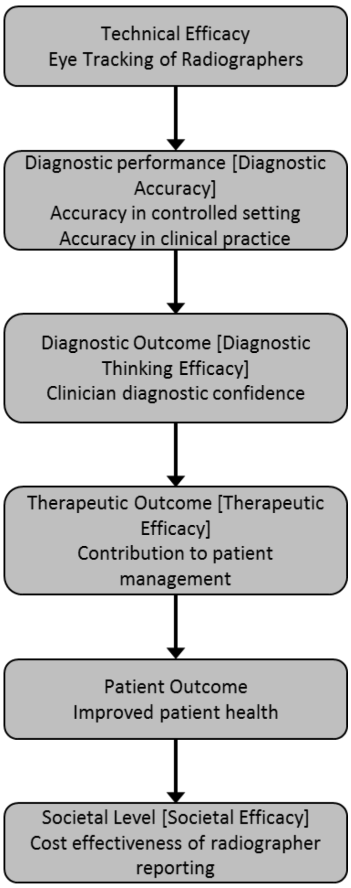


Figure 2.3 Brealey's hierarchy of efficacy for radiographer reporting with possible outcome measures for each level (adapted from Brealey 2001[62] and Brealey and Scally 2008[63])

Technical efficacy, level one on the hierarchy, examines the technical aspects of the new technology. For new techniques or modalities this is often the basic science that underpins the investigation, such as contrast resolution or tissue enhancement.[57] The equivalent of this for radiographer reporting is the visual search strategies and the perceptual-cognitive process of image interpretation.[179-181] The second level of the hierarchy, diagnostic accuracy efficacy, incorporates both the technology and the observer interpreting the images. Measures of diagnostic accuracy include accuracy, yield, error rate, sensitivity, specificity, positive and negative predictive values and receiver operator characteristic (ROC) curves.[182 183] A key concept contained within the framework is the separation of diagnostic accuracy and diagnostic performance within the second hierarchical level. This delineation recognises the difference in performance that occurs between a test environment (diagnostic accuracy) and clinical practice (diagnostic performance).[62]

Level 3 of the hierarchy is diagnostic thinking efficacy/diagnostic impact (DI) and is the impact that imaging reports have on the diagnostic decision-making of clinicians.[184] Diagnostic impact, an empirical proxy to patient outcome, determines the influence that radiology investigations have at the point closest to the test, thus minimising confounding variables. First used in the clinical evaluation of computed tomography of the head,[56] diagnostic impact has become an established intermediate outcome measure.[184] Research has also been performed which examined the influence of radiographer skeletal reports on emergency physician and general practitioner management plans, level 4 therapeutic impact/efficacy, with no statistically significant difference found in this small structured assessment.[185] Radiographer reporting is in a similar position to radiology, with a paucity of research that has examined the higher levels of efficacy. A notable exception, as mentioned previously, is the recent randomised control trial which examined the impact on patient outcome (level 5 efficacy) and cost effectiveness (level 6 efficacy) of immediate reporting of trauma skeletal X-rays by reporting radiographers.[39 186]

A fundamental assumption used within the framework proposed by Brealey and Scally is that the performance of a radiologist is taken to be the reference standard.[62 63] This has profound implications, as it has been well documented that there is significant variation in the accuracy of radiologist interpretations.[187-189] This emphasises the importance of a robust reference standard, as discussed in Chapter 2.7, and was considered when designing the study protocol (Chapter 3.2.7).

2.7 Use of chest X-rays in health screening and variability of chest X-ray interpretation accuracy

Chest X-rays (CXRs) have long been used as a tool to assist diagnostic and therapeutic decisions. Despite the development of new imaging modalities and techniques, the chest X-ray (CXR) remains central to many patient pathways. In 2013-14 there were 7.4 million chest X-rays performed in the NHS in England.[190] The use of chest X-rays has evolved but despite its widespread use, considerable variation exists in the interpretation of them. A critical review of the literature has been performed. Particular emphasis has been placed on studies that have utilised the receiver operator characteristic curve (ROC) and alternate free response ROC (AFROC) methodologies. A detailed discussion of literature has occurred in Chapter 2.7 below and related to the results of the current study in Chapter 5.4. Summary tables are included in Appendices 1 and 2.

The mass survey chest X-ray is a classic example of an investigation that had a relatively high impact on clinicians' diagnostic confidence, yet produced a very small impact on therapeutic management due to the lack of effective drug therapy.[191] In addition, it is often argued that the confidence placed in the ability for a chest X-ray to assist in the diagnosis of tuberculosis was often misguided, mainly due to the significant observer variation that can occur during interpretation, even by highly skilled and trained observers.[68 71 74]

A wealth of studies in the literature have assessed the effectiveness of the chest X-ray in screening programs, especially tuberculosis.[66 69 71 72 154 174 192-202] This includes a number of works utilising mortality as an outcome measure to determine the impact of screening chest X-rays on patient health.[155 174 200] The primary goal of a screening program is to detect sub-clinical disease in an asymptomatic population.[193 199] While this goal was common to both military and public health mass chest X-ray programs, the ultimate desired outcome was fundamentally different. Excluding those recruits which demonstrated evidence of tuberculosis from service was the explicit aim of the military screening program implemented by many governments prior to, and during, World War 2.[196] Once a chest X-ray was determined to show signs of active or previous infection, the candidate was barred from enlisting in the armed services and discharged back into civilian life, often without provision for adequate referral for treatment.[196] The public health perspective, in which the primary concern was reducing the mortality and morbidity of tuberculosis through case detection, isolation and treatment, contrasted with this. In this scenario, it was proposed that early detection would result in simpler treatment, improved patient outcome and, in certain circumstances, a decreased risk of contagion.[69 155 174 193 197 200 201 203] The narrow focus on excluding recruits with suspected tuberculosis while ignoring the opportunity to treat disease and improve health is highlighted by Haygood as a flaw in the military screening program, and a missed opportunity to control this devastating disease and limit the impact on society.[196]

The retrospective analysis of such a vast number of chest X-ray examinations provided researchers with invaluable data. A crucial area examined was the variation that occurred when interpreting chest radiographs, with the early work of Hodges[66] and Freer[194] highlighting not only the significant inter-observer, but also intra-observer variation that occurs when interpreting radiographs. Initially driven by the differences in rejection rates from different screening stations during army recruitment, this variability in interpretation was emphasised by the contribution that recruits from different enrolment stations made to patients subsequently treated at the US Army's hospital for tuberculosis. Retrospective analysis conducted by Long and Stearns (1943), reported in Freer,[194] demonstrated that a disproportionately high number of patients originated from only a few screening centres, which prompted review of the initial chest X-rays. Of the 100 chest X-rays initially reported as positive randomly selected by Freer for re-interpretation, definite evidence of tuberculosis was demonstrated in only 66 cases.[194] The work of Myers (1946), reported in the review by Holman (1950), concluded that between 10% and 15% of all recruits rejected from military service due to tuberculosis had chest X-ray findings consistent with acute non-tuberculosis infection.[197]

This high discrepancy level is almost unanimously reported across the literature when examining screening chest X-ray interpretation. Clark *et al.* found that of the 1,577 patients admitted to hospital with tuberculosis, 12% had non-tuberculosis disease and 38 demonstrated no pulmonary disease at all.[70] The work of Hodges reported that only 51 of the 69 patients identified in the screening program of 7,841 chest X-rays had the abnormality confirmed on follow up,[66] while Carroll *et al.* established the initial diagnosis of tuberculosis or other pathology in only 86.5% and 55.6% of cases respectively.[192] In their analysis of a lung cancer screening program, Lilienfeld and Kordan established a combined inter-observer variation rate of 67% once technically inadequate images were excluded.[71] Of interest was the sub-analysis conducted in this study, with good (89%-92%) agreement found when the chest X-ray was interpreted as normal, but only moderate agreement was found when the chest X-ray was reported as abnormal, concluding that disease prevalence influences the agreement rates between observers.

More recent work into observer variation when reporting chest X-rays for tuberculosis have reported mixed findings. Analysing the historic data, this significant variation was confirmed in the recent analysis of Koppaka and Bock for the World Health Organisation (WHO),[204] who identified rates of over (false positive) and under (false negative) reading of screening chest X-rays for tuberculosis ranging between 0.3%-1.7% and 21%-39% respectively. The work conducted by Balabanova and colleagues examined the variability in chest X-ray interpretation of Russian and British medical practitioners.[75] When reporting a bank of selected chest X-ray images, only fair (Kappa K=0.38) agreement was found for combined overall performance, but they found different results when they examined the performance of different professional groups. Consultant radiologists (K=0.497) outperformed both TB specialists (K=0.368) and respiratory physicians (K=0.284) when agreement for normal and abnormal was investigated, a trend that was replicated when analysis was repeated for findings that were consistent with tuberculosis (K=0.448, 0.377, 0.386). These findings were in contrast to the study of Abubakar *et al.* who reported mixed results, finding higher agreement in cases demonstrating tuberculosis between physicians (K=0.64) than radiologists (K=0.54), although this did not reach statistical significance.[74] They did however find a significant difference in agreement when interpreting normal examinations, with radiologists outperforming physicians in this sub-analysis (K=0.84 vs. 0.46).[74] Given that many of the observers used to interpret screening TB chest X-rays for both military and public health surveys were non-radiologist physicians these different levels of variability need to be considered.

Variable results were reported in the study of Zellweger *et al.* who found fair (K=0.55) overall agreement for the three observers when interpreting a random selection of screening chest X-rays from high-risk immigrants in Switzerland.[76] When the experience of the observers was considered, they found good (K=0.84) overall agreement on the presence of active TB. These findings need to be taken in context, as the third observer was a junior medical officer, whose performance was being compared to the image interpretation ability of senior respiratory physicians.[76] Experience was also found to be a factor in the analysis of Abubakar *et al.* who reported improved image interpretation accuracy with increased years in speciality and the number of cases of TB seen in practice per year.[74]

Even though variation in chest X-ray interpretation is well established, this was ignored in the recent work of Eisenberg and Pollock when determining the yield of pre-employment screening chest X-rays in employees who had a positive tuberculin skin test (TST).[205] By utilising the reference standard (expert consultant radiologist review) in only those cases initially interpreted as abnormal by the initial reading radiologist, they potentially missed those cases that were incorrectly read as normal in the first instance. This may have introduced work-up bias into the study,[206] although the reference standard did not contradict any of the initial readings in the 159 abnormal chest X-rays examined.

The accidental inclusion of a small number of chest X-rays that were reported twice by the same observer in the study of Lilienfeld and Kordan allowed intra-observer variation to be analysed. [71] They found that an observer was no more likely to agree with their own diagnosis than with another observer.[71] Significant intra-observer variation has been reported in several other studies, including the work of Garland in 1949[197] and 1959[204] although this intra-observer variation was always less than inter-observer variation, with levels 8% and 19%-24% compared to 16% and 27%-30% respectively. When re-reporting X-rays 24 hours later, intra-observer agreement between a group of five radiologists demonstrated variability, ranging from moderate to substantial (K=0.33 – 0.88), although accuracy did increase slightly.[207] The analysis conducted by Zellweger *et al*. report contrasting results, finding good and very good intra-observer agreement when experienced respiratory physicians interpreted screening chest X-rays for the presence of active tuberculosis, although this was significantly better than the less experienced observer, who only had a fair level of agreement.[76] Consequently intra-observer variability was considered when designing the study.

Several causes are postulated for the variation in accuracy when interpreting screening chest X-rays for tuberculosis. The significant contribution made by the technical specifications of the imaging systems must be acknowledged, with the majority of the mass survey chest X-ray programs utilising either miniature chest X-rays or photofluoroscopic images for interpretation. While these developments and subsequent reduction in the cost of an examination made the application of mass screening feasible,[191 196] the inherent limitations must be recognised. The ability of an observer to detect subtle yet important findings on a smaller image when compared to a standard 35 cm by 43 cm X-ray film will be reduced, with the work of Wagner (1999) recognising that maximum observer performance is limited by the physical characteristics of an imaging system.[160] The work of Lorimer suggests that it was these technical factors rather than a failure to observe that led to many of the errors in diagnosis.[65] These limitations were recognised in the study of Carroll *et al.,* with suspicious cases detected on miniature chest X-ray recalled for examination with a dedicated 35 cm by 43 cm (14 x 17 inch) chest X-ray, confirming tuberculosis infection in 64 of the 74 cases re-examined (86.5%).[192] The recent work of Balabanova *et al.* that examined the ability of chest X-ray interpretation of different Russian medical specialities made use of non-medical specification image software, a factor that must be considered when considering the performance of the observers reported in their trial.[75]

The characteristics of the observers involved in image interpretation of screening chest X-rays have also been examined as contributing to the high discrepancy rates.[69] Factors including carelessness, poor search strategy, monotony of the task and minimal skill of the interpreters are all proposed to play a role, but they found mistakes were also made by very experienced observers.[194 197] Examination of the characteristics of the medical officers responsible for interpreting these radiographs showed that they were often inexperienced, with little formal training in image interpretation. This lack of specialist training is highlighted by the work of

Brodeur, who reported training programs for medical staff ranging in length from a few days to eight weeks, and outlined his own curriculum that used five or six images when demonstrating and teaching normal anatomy and radiographic appearances.[208] Screening mammogram interpretation has similar characteristics to screening chest X-rays for tuberculosis; high volume, lower disease prevalence and monotony of task. In contrast to the results of tuberculosis screening, recent work that examined observer fatigue in screening mammogram reading found that the order in which expert mammographers read the cases did not influence accuracy.[209] This most likely reflects on the robust and rigorous training that practitioners undergo prior to interpreting mammograms when compared to the variable education given prior to screening chest X-ray reading.

It is important to note that the primary aim of many of these training programs was to train observers to highlight radiographs which required further investigation, rather than to provide a radiological diagnosis of the findings.[199] In keeping with the nature of the screening program, a heavy emphasis was placed on the appearances of tuberculosis, as well as other common and significant non-tuberculosis pathologies.[208] This variation in specialist training in interpreting one of the most complex radiological plain film examination go some way to explaining the differences seen in screening chest X-ray interpretation.

The impact on patient outcome of a mass survey chest X-ray programme was examined in the work of Enterline and Kordan[174] and Wylie,[155] who both found that screening for tuberculosis with a chest X-ray decreased mortality. Wylie pursued a more conventional approach, examining the five year mortality for patients who had participated in a multi-phasic screening program, concluding that chest X-ray screening for TB was highly efficient, although the relatively small sample size and few deaths within the cohort limit these findings.[155] Enterline and Kordan however employed a more novel approach by exploiting the inherent error in image interpretation.[174] Over 200,000 chest X-rays were re-reported by non-radiologist physicians and a control group formed from the proportion of patients whose chest X-rays were initially incorrectly interpreted as normal. The use of an unintentional initial incorrect (normal) diagnosis, identified at subsequent second reading, as the intervention arm did not withhold relevant and important information from the clinical team. Patient outcomes could be compared for the prompt (initial correct abnormal) and delayed (incorrect initial normal) chest X-ray reports and the impact of incorrect radiology interpretation quantified. They reported 97 deaths from tuberculosis within the cohort, 43 within the intervention group and 54 in the control group with the incorrect initial interpretation that did not receive prompt treatment, a statistically significant difference. [174]

Although partly explained by the wide variation in observer accuracy and the different imaging systems utilised within the different screening programs, a significant contributor to the lack of improved patient outcome was the lack of an effective treatment, not only for tuberculosis but also for lung cancer. Key milestones in the medical management of tuberculosis occurred during the period of mass survey chest X-rays, beginning with the discovery of streptomycin in 1944, isoniazid in 1952 and rifampicin in the late 1960s.[210] This resulted in markedly improved patient outcomes, mortality falling dramatically from tuberculosis being the leading cause of death in young men in 1940.[196] This is contrasted by the results of the lung cancer screening programs in the 1970s. Many studies demonstrated an increased rate of detection and survival but no change in mortality, but there was no direct analysis between having and withholding chest X-rays in these initial trials.[133]

Survival rates for lung cancer have shown minimal change over the last 40 years. There has only been a modest increase in survival time, 9.5% of patients survive ten years from diagnosis in 2010-11 compared to 4.6% in 1971-2, an important factor given the worldwide and UK burden of this disease.[98] Recent analysis of cancer survival by type still confirms lung cancer has poorer survival compared to breast and colon,[99] and it is suggested that the late stage at diagnosis is a contributing factor.[98] Focus is being placed on early and accurate diagnosis for cancer, particularly lung cancer, in an attempt to improve patient outcomes. Effective screening programmes for breast and colon cancer may contribute to the earlier diagnosis and outcomes associated with these diseases. Although chest X-rays for lung cancer screening were found to be ineffective,[73 211] more recent studies that examined the use of low dose chest CT have shown a reduction in patient mortality.[139 212 213]

Variation between consultant radiologists has also been examined outside of a screening setting. The 1999 study by Robinson et al. found three experienced consultant radiologists only agreed in 151 of 205 (74%; K=0.68) skeletal X-rays, 61 of 100 (61%; K=0.50) chest X-rays and 50 of 97 (51%; K=0.42) abdominal X-rays.[188] These findings correlate with the work of Tudor, who reported moderate agreement between five radiologists interpreting a bank of 50 X-rays with the clinical history (K=0.58).[214] These results were broadly comparable to the intra-observer agreement between radiologists when re-reporting the same bank of X-rays 24 hours later, Kappa between 0.33 and 0.88.[188 207]

Image interpretation by radiographers is not a new concept. Swinburne (1971) is credited with first raising the idea in the UK that pattern recognition could be performed by trained radiographers in order to fast track suspicious X-rays for an urgent radiologist report,[215] although the concept had been examined in the context of lung cancer screening by Sheft *et al.* previously.[33] Harnessing the skills of the entire clinical team, Berman and colleagues (1985) demonstrated that an abnormality detection system, the 'red dot', that utilised the image interpretation skills of radiographers in a trauma setting decreased the diagnostic errors made by emergency clinicians.[216] Several studies have examined the ability of radiographers to detect abnormalities on skeletal X-rays,[217-223] and have established that radiographers possess the core image interpretation skills to identify abnormal trauma X-rays.

The abnormality detection skills of radiographers have also been used in other areas. In order to maximise the detection rate and avoid potentially life-threatening omissions, it has been suggested that screening X-rays receive dual interpretation, both for chest images and mammograms as well although this has not been universally accepted or implemented in practice.[71 200 224] The burden of such a high volume of images requiring interpretation was significant, and one potential solution to this problem proposed by Sheft *et al.*[33] and Flehinger *et al.*[34] was to utilise specially trained radiographers in an image interpretation role. Both of these controversial works demonstrated that specially trained radiographers were able to distinguish between normal and suspicious or abnormal chest X-ray films to a level comparable with consultant radiologists. When first proposed by Sheft *et al*. [33] this idea was criticised in the radiological literature, with suggestions that "anyone with even ordinary ability would be expected to score passably well in such a loose test system" (p.76), and that most radiologists would be unwilling to participate in a screening program that involved image interpretation by radiographers.[225]

Considerable work has examined the diagnostic accuracy of trained reporting radiographers when interpreting skeletal X-rays. These analyses have been performed in both academic and clinical practice. A review of nearly 7,000 skeletal cases interpreted in an objective structured examination as part of an accredited postgraduate training programme, Piper *et al*. found high levels of performance, with average sensitivity 91.6% - 96.7%, specificity 92.1% - 94% and accuracy 92.5% - 93.9%.[23] Using patient re-attendance to the emergency department of a single hospital, Robinson *et al*. found discordant radiographer reports in only 29 of the 1,130 cases who presented from a consecutive series of more than 11,000.[38] These findings correlate well with the multisite clinical trial conducted by Piper *et al*. which reported radiographer accuracy between 97.1% - 99.8% for 7,179 examinations.[22] Typically, when diagnostic accuracy results from structured assessment and clinical practice are compared, sensitivity is frequently lower. Structured assessments tend to utilise an enhanced proportion of abnormal cases. The reasons for a greater number of abnormal cases is two-fold. A balanced design (equal proportion of normal and abnormal cases) is the most efficient methodology, in terms of number of cases and number of observers, to discriminate between observer accuracy. A greater number of abnormal cases also facilitates the inclusion of difficult, subtle and infrequent but important pathologies. The reduced sensitivity reported in clinical practice is often due to the lower disease prevalence (abnormal cases) encountered in cases drawn directly from a routine workload. Piper *et al.* found the converse. The mean sensitivity reported for the structured assessment (50% prevalence) was between 90.3% - 91.7% for the three cohorts. In comparison the mean sensitivity of the reporting radiographers in clinical practice, with a mean prevalence of 14.9%, was 97.6% (95%CI 96.5 – 98.6%). The systematic review performed by Brealey *et al.* provided the conclusive evidence.[24] On review of 28,900 skeletal examinations, reporting radiographers were found to have an average sensitivity and specificity of 92.6% (95%CI 92% - 93.2%) and 97.7% (95%CI 97.5 – 97.9%) respectively. Recent work has established that not only is immediate reporting of skeletal X-rays by reporting radiographers from the emergency department accurate,[186] it is also cost effective.[141]

When compared to the extensive research base underpinning skeletal X-ray reporting by radiographers, there is a relative paucity of studies that have explored chest X-ray interpretation by radiographers. Following on from the work examining the ability of trained radiographers to identify X-rays suspicious for lung cancer,[33 34] the next study to examine radiographer abnormality detection for chest X-rays was Sonnex et al.[226] Their six month evaluation at a specialist hospital included 8,614 consecutive chest X-rays and 17 radiographers. The radiographers had received in-house training from consultant radiologists on chest X-ray interpretation, with a focus on postoperative appearances following cardiothoracic surgery, as well as regular feedback on performance. Radiographer sensitivity and specificity for the detection of abnormalities that required urgent intervention was 90% and 99% respectively.

Trained reporting radiographers have been included as observers in eye tracking analyses. These studies have examined nodule detection rather than the provision of a clinical report,[81] [227] and reporting radiographers are often pooled with consultant radiologists when comparing novice and expert observers.[228] Two studies have examined the performance of trained reporting radiographers providing clinical reports for chest X-rays. Four thousand chest X-rays, with normal cases and a range of pathologies, were interpreted by 40 radiographers at the end of an accredited postgraduate training programme with 95.4% sensitivity, 95.9% specificity and 89% agreement.[41] The image test bank used by Piper *et al.*,[41] with a higher disease prevalence than is typical in clinical practice, may have influenced participant performance. The results are an important first step, as they found reporting radiographers are accurate practitioners prior to reporting in clinical practice. At present, only one study has examined the accuracy of reporting radiographer chest X-ray reporting in clinical practice, a clinical audit of practice. The study was a review of a random sample of reports provided by a single reporting radiographer who had completed accredited postgraduate training in chest X-ray reporting and was accredited as an advanced practitioner with the College of Radiographers. The retrospective study required three consultant radiologists to review the radiographer chest X-ray reports and corresponding images. High concordance was found with the reviewing consultant radiologists (agreement 92% - 96%; Kappa 0.83 – 0.91) and the reporting radiographer in the 99 cases included for review.[42] This small study confirmed that the reporting radiographer was a safe and effective practitioner when reporting adult chest X-rays within a supportive multidisciplinary team.

2.8 Summary of existing research and gaps in the current evidence

Chest X-rays form a key role in many patient pathways, used by clinicians to guide diagnostic and therapeutic decisions. A considerable body of evidence, both historic and contemporary, demonstrates significant differences in the accuracy of chest X-ray interpretation. Variability is found regardless of the professional background of the reader; consultant radiologist,[73 77] non-radiology consultant physician,[74 82] trainee radiologist,[229] junior medical staff[230] and radiographer.[33] The bulk of the literature has focused on chest X-ray interpretation by medical practitioners. A small number of studies have examined chest X-ray interpretation by radiographers. Early work examined the ability of radiographers with some additional training to highlight abnormal chest X-rays for urgent radiologist reporting, with promising results.[34 226] Other studies have examined radiographer accuracy in nodule detection, also with good results.[227 228] A common limitation of these research designs is that the radiographers were not required to arrive at a diagnosis or provide a definitive clinical report.

Few studies have examined the diagnostic accuracy of radiographer clinical reporting of chest X-rays. The work of Piper *et al*. found high sensitivity and specificity when a cohort of radiographers completed an objective structured examination at the end of an accredited postgraduate training programme.[41] Although conducted at the end of a training programme, the radiographers were not yet reporting chest X-rays in clinical practice. Accuracy may improve over time as radiographers gained clinical experience,[231] or performance could decline with time following the cessation of training.[232] Another limitation is that no direct comparison was made with consultant radiologists, the established experts in medical image interpretation and the benchmark on which performance is measured. The clinical audit conducted by Woznitza *et al*. reported agreement between a single reporting radiographer and three consultant radiologists for a random sample of chest X-rays reported in clinical practice.[42]

Although susceptible to verification bias, they found comparable radiographer-radiologist and inter-radiologist agreement.

Clinicians' use imaging investigations to assist decision-making. A hierarchy of efficacy has been designed that evaluates the role of imaging investigations, and how clinicians integrate clinical reports.[57] The majority of the existing literature has examined cross-sectional imaging, CT[56 172] and MRI.[158 163] Little research has investigated the role of chest X-rays in clinician decision-making,[233] and even less the impact of a radiographer report.[234]

The majority of reporting radiographers in clinical practice report skeletal X-rays,[35] and a considerable evidence base supports this.[23 24] However, as an increasing number of radiographers' complete accredited postgraduate training in chest X-ray reporting, further research is required to support practice. Firstly, research is needed to establish the diagnostic accuracy of chest X-ray interpretation by reporting radiographers. Secondly, chest X-ray reports provided by reporting radiographers need to be usable to clinicians; they need to have a positive influence on diagnostic decision-making. In line with Royal College of Radiologists and College of Radiographers guidance, the performance of chest X-ray reporting radiographers needs to be comparable to consultant radiologists.[44] At present, there is a paucity of research that has examined these questions.

Chapter 3 - Methodology

3.0 Overview of methodology used in the study

The study reported in this thesis is a quantitative assessment of the diagnostic accuracy of consultant radiologist and reporting radiographer chest X-ray interpretation, and the influence that these reports have on clinicians' diagnostic decision-making. The study was a quasi-experimental design where reporting practitioners (consultant radiologists and reporting radiographers) interpreted a bank of chest X-rays with performance measured against a robust reference standard diagnosis. Subsequently the influence of chest X-ray reports on clinicians' diagnostic decision-making was explored using a before and after approach, with changes in diagnosis and confidence levels in the diagnosis established. The conceptual framework for the current study and justification of the methodological approach are provided in Appendix 12.

3.1 Research governance and ethics

All research and audit conducted as part of this study complied with the principles for ethical medical research,[235] Health Research Authority standards of practice, Good Clinical Practice Guidance,[236] and was in line with Data Protection requirements.[237]

The audit performed to determine the prevalence of diseases associated with chest X-rays at the clinical site where cases were selected (Chapter 3.2.3 and Appendix 3) was registered with the Trust clinical quality department.

Health Research Authority approval for the diagnostic accuracy and diagnostic decision-making study was gained prior to the research commencing from the National Research Ethics Service (Appendix 4). Research and Development approval was received from the host institution prior to the study commencing (Appendix 5). The study was funded by a grant from the College of

Radiographers Industry Partnership Scheme (CoRIPS) and was registered with the National Institute for Health Research Clinical Research Network portfolio.

Chest X-rays used in the study were taken from a consecutive series performed for clinical reasons at the Trust; no additional radiation exposure was required for the study. All cases included in the study were pseudoanonymized by the radiology PACS team, both the randomised, consecutive sample and for cases taken from the radiology department discrepancy meetings; no patient identifiable information was included. The original clinical reports provided at the time of the examination of each patient were compared to the reports provided by the expert chest radiologists for the purposes of this study. No significant clinical abnormalities were found to have been missed. Additionally, the case note review performed by the professor of medicine, again for the purposes of this study, confirmed that no patient management decisions should have been different.

Participants in the research project were NHS employed consultant radiologists, reporting radiographers, medical doctors and an academic radiographer. They were given participant information sheets relevant to their role (expert radiologist, arbiter, diagnostic accuracy participant, clinician for diagnostic influence study) which explained the requirements for participation in the study (Appendices 6 – 9). Informed consent was obtained from all participants prior to data collection, and participants were free to withdraw at any time (Appendices 10 and 11). All participants were assigned a unique study identifier that ensured that all information collected as part of the diagnostic accuracy and diagnostic decision-making study was anonymised. Study participants only had access to the pseudoanonymised patient details; no patient identifiable information was included on the X-rays or the case summaries. A small honorarium was paid to participants in acknowledgement of the time taken to complete the study.

3.2 Part 1 – investigation of diagnostic accuracy

Primary Research Question Is the accuracy with which a group of consultant radiologists interprets a bank of chest X-rays from adult hospital based patients comparable to the accuracy of a group of reporting radiographers?

Incorporating the suggestions of Tourassi[238] and the CONSORT extension for non-inferiority designs,[48] and considering the need to distinguish between clinically and statistically significant findings[129]:

Alternative Hypothesis [non-inferiority approach, (Appendix 12.6)] That there is no clinically significant difference in the accuracy of adult chest X-ray interpretation between consultant radiologists and reporting radiographers.

Null Hypothesis [non-inferiority approach] That the accuracy of adult chest X-ray interpretation of consultant radiologists will be significantly (both clinically and statistically) superior when compared to the accuracy of reporting radiographers.

There is no established benchmark of performance that is required for practitioners to interpret chest X-rays in clinical practice. The study is therefore not interested in the absolute performance of the reporting radiographers relative to a reference standard diagnosis, which in the current study are cases in which two expert chest radiologists are in agreement, rather the relative performance of the cohort of consultant radiologists and reporting radiographers. This is in line with recent guidance on radiographer reporting, which suggests that reporting radiographers must perform at a level comparable to consultant radiologists.[44 239] In order to achieve this comparison, the diagnostic accuracy of the consultant radiologists and reporting radiographers will be indirectly compared to the reference standard diagnosis. For the purpose of this study, hospital based patients comprised of patients referred for a chest X-ray via the emergency department, outpatient or as an inpatient. This was a pragmatic decision; medical notes were required to establish the final clinico-radiological diagnosis, and obtaining medical records from general practice would have been too complex to achieve within the timeframe of the study. Chest X-rays are one of the most frequently performed imaging investigations in England.[94] As a high volume test, chest X-rays play a role in many patient pathways and require considerable resource to provide a clinical report.

Secondary Research Question What is the level of agreement between the expert consultant chest radiologists used to establish the reference standard diagnosis for the image bank?

Agreement between consultant radiologists in chest X-ray interpretation in the literature is variable.[74 76 188 207 240] The anticipated levels of agreement reported by Robinson *et al*. (Kappa=0.5)[188] and Tudor *et al.* (Kappa=0.33 – 0.88),[214] were used for sample size estimates for the number of cases which would be needed when constructing the image bank based on expert consensus.

Alternative Hypothesis [superiority approach] That there will be a high level of agreement between the expert chest consultant radiologists when interpreting chest X-rays.

Null Hypothesis [superiority approach] That there will be a low level of agreement between the expert chest consultant radiologists when interpreting chest X-rays.

3.2.1 Methodology

A quasi-experimental assessment of diagnostic accuracy was employed, using a free-response paradigm and a non-inferiority approach. This examined level 2 efficacy within the framework and is illustrated in Figure 3.1.[57]

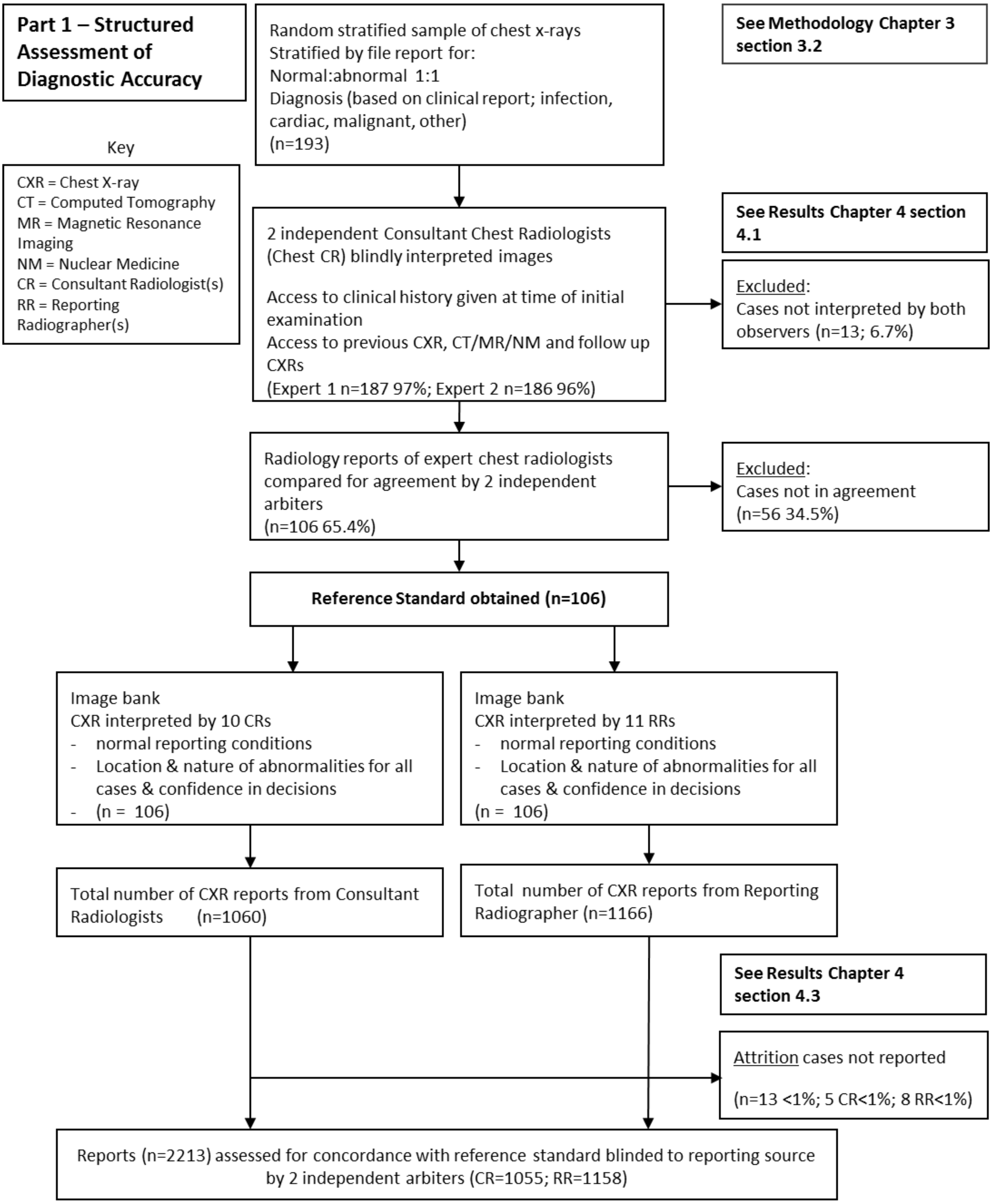


Figure 3.1 Flow diagram of diagnostic accuracy (part 1) study including exclusions and attrition

3.2.2 Case Selection and Allocation

A consecutive series of adult chest X-rays (CXRs) performed for clinical reasons at the Trust (Chapter 3.1) between 1st April 2011 and 30th March 12 were retrieved from the electronic patient record by the IT department. An automated report was generated, using the standardised search terms detailed in Appendix 13.

Inclusion criteria for the study were:

- Adult patient (>16 years)
- Chest X-ray examination
- X-ray performed when the patient was referred by a hospital based clinician:
    - Emergency Department (ED)
    - In-patient (IP)
    - Out-Patient (OP)

Exclusion criteria for the study were:

- Paediatric patient (<16 years)
- Any other imaging modality (ultrasound, fluoroscopy, computed tomography, magnetic resonance imaging)
- All other X-ray examinations
- Examinations referred by a non-hospital based clinician:
    - General practitioner (GP)
    - Sexual Health (DOSH)
- Previous chest X-ray included in the study.

Chest X-rays referred from general practice and sexual health were excluded from analysis due to the logistical problems associated with case note retrieval from remote clinical sites.

Examinations that met the inclusion criteria were sorted into descending chronological order. The PACS team in Radiology assigned a unique case identification number (pseudoanonymised). The random number function of Microsoft Excel 2010 was used to generate the cases selected for inclusion in the study. Initial stratification of the pseudoanonymised cases (normal and for each disease category) was performed by the chief investigator, based on the clinical report provided by the reporting practitioner at time of clinical interpretation. Initial stratification maximised the efficiency and workload of the expert chest radiologists. This initial stratification did not form part of the classification system for determining the inclusion or exclusion of a case.

In order to determine diagnostic accuracy using the free-response approach, a robust reference standard diagnosis, or ground truth, is required for all cases.[64 241] As such, cases in which both expert chest consultant radiologists were in agreement were included in the image bank for interpretation by the participant reporting radiographers and consultant radiologists. The image bank was stratified for a normal:abnormal ratio of 1:1 and for a disease category (infection: cardiac: malignancy: other) ratio of 3:3:1:3 according to the interpretation of the expert chest radiologists. A balanced design, with an equal proportion of normal and abnormal cases, is the most efficient for diagnostic accuracy studies as it allows small differences in observer performance to be detected for a given number of cases and observers.[242] The proportion of cases within each disease category was based on the proportions from an audit of most frequent discharge diagnoses associated with a chest X-ray performed at the study site (see Appendix 3). A small proportion of subtle normal and difficult abnormal cases were included for the normal cases and for each disease category. These were drawn from the monthly radiology department discrepancy meeting and the teaching files of the chief investigator. Inclusion of subtle cases ensured that a sufficient range of true positive and false positive responses were generated by the participants and ensured the data was appropriate for analysis.[243 244]

3.2.3 Constructing a clinically representative image bank – audit data

An audit was performed which established the disease prevalence at the recruiting department. The medical records of patients who had had a chest X-ray and who were discharged from the emergency, outpatient and inpatient services over a 12 month period were reviewed for the primary coded disease at discharge. The project was deemed an audit of clinical practice and therefore ethical approval was not required. The audit was registered with the Quality and Audit department in line with best practice recommendations. These diagnoses were reviewed and collated into broad disease categories by a professor of medicine. Forty percent of cases were associated with a pathological diagnosis (40% disease prevalence). The most common broad diagnostic categories were infection and cardiac/pulmonary oedema which each accounted for approximately one third of all X-rays performed (Appendix 3). Malignancy was demonstrated in approximately 7% (n=7) of abnormal cases. Comparison with national disease datasets revealed that these findings were broadly comparable to population figures available for England.[245] A phased approach to image interpretation was used to ensure that the final image bank aligned with the required proportions (53 normal; 20 infection, 12 cardiac, 7 malignant, 14 other).

3.2.4 Sample size calculation – diagnostic accuracy study

3.2.4.1 Primary analysis – Accuracy of consultant radiologists and reporting radiographers

Adopting a free-response methodology, jack-knife alternate free-response receiver operator characteristic (JAFROC) curves were used (Appendix 12.4.1.2).[246 247] To detect an effect size (difference between consultant radiologists and reporting radiographers) of 0.08 in the Figure of Merit, with 10 observers in each group and a normal/abnormal CXR ratio of 1:1, 105 cases were required for the test bank, accepting a 20% chance of a Type II error and a 5% possibility of a Type I error.[241] To detect a 10% difference in the area under the ROC curve and assuming 10 observers in each group, and a normal/abnormal CXR ratio of 1:1, 101 cases were required for the test bank, accepting a 20% chance of a Type II error and a 5% possibility of a Type I error.[242] This powered the study to detect a small difference between highly accurate observers, if one existed. The results of both Chakraborty (JAFROC) and Obuchowski (ROC) compare well; a sample size of 106 was used, to allow for a 1:1 ratio of normal:abnormal cases.

3.2.4.2 Secondary analysis – Agreement between expert consultant chest radiologists

A staged approach was used to establish the reference standard diagnosis to ensure that the stratified sample was achieved. It was estimated that a total of 220 chest X-rays would need to be interpreted by the two expert chest consultant radiologists to obtain the 106 cases required for the image test bank. The estimate was based on anticipated levels of expert agreement reported in the literature.[74 76 188 207 240] Although not a primary analysis, the sample size calculations of Sim and Wright indicate that for a prevalence of 50% and to detect a Kappa of K=0.4 a total of 66 cases were required to be interpreted by each observer using a 2-tailed test with 5% possibility of a type I error and 80% power.[248]

3.2.5 Study participants – diagnostic accuracy study

Interpretation of the images was performed by trained professionals. Participants (expert chest consultant radiologists, consultant radiologists, reporting radiographers, and arbiters) were voluntarily recruited through convenience and snowball sampling. An insufficient number of reporting radiographers was recruited through convenience sampling, so an advertisement was placed in the radiography professional magazines *Synergy: Imaging and Therapy Practice* and *Synergy News* (Appendix 14). Each participant was provided with a participant information sheet appropriate for the task being performed (Appendices 6 – 8) and provided signed informed consent to the chief investigator (Appendix 10). All participants were assigned a unique study reference by the PACS team to ensure anonymity and were free to withdraw at any time. A small honorarium and travel expenses were paid to acknowledge the time and effort required to participate, in line with research governance and with ethical approval (Chapter 3.1).

General demographic details were recorded for each participant. These consisted of:

- Profession (Consultant Chest Radiologist/Reporting Radiographer/Consultant Radiologist/Academic Radiographer)
- Experience – years post qualification (Reporting Radiographers = obtaining an adult chest X-ray reporting qualification, Consultant Radiologists = obtaining fellowship of the Royal College of Radiologists)
- Practice – number of chest X-ray reporting sessions per week, number of cases reported per year

For the purposes of this study, the expert chest consultant radiologists were staff that held a consultant post within the NHS, specialist registration in clinical radiology with the General Medical Council and a specialist interest in chest/thoracic imaging. The consultant radiologists were staff that held a consultant post within the NHS and specialist registration in clinical radiology with the General Medical Council. The reporting radiographers were consultant or advanced practitioner radiographers that held registration with the Health and Care Professions Council and Masters level qualification (MSc, postgraduate diploma, postgraduate certificate) in adult chest X-ray interpretation accredited by the College of Radiographers. The arbiters were a senior academic radiographer that holds registration with the Health and Care Professions Council and doctoral level qualification and a professor of radiology that held a consultant post within the NHS and specialist registration in clinical radiology with the General Medical Council.

3.2.6 Image interpretation

3.2.6.1 General considerations

All image interpretation occurred under normal viewing conditions. Digital Imaging and Communications in Medicine (DICOM) images were viewed on paired high resolution Picture Archive and Communication System (PACS) workstations in observer controlled lighting conditions free from distractions and background noise.

Participants entered the report for each case directly into an electronic proforma (Microsoft Word 2010) by either directly typing the report or using voice recognition software (Dragon Natural Speaking, Nuance). Training, which consisted of three sample cases, was given to all participants by the chief investigator. This allowed familiarity with the image interpretation task, the hardware and software used, the voice recognition tool and the proforma. In addition, to provide the data for JAFROC analysis, the expert chest radiologists were shown how to assign a location to each abnormality on the image and construct the abnormality list for each image. Free response analysis relies on the participant locating abnormalities on an image and assigning a rating for each abnormality based on their confidence level (scale 1-4; 1 = uncertain, 4=definitely abnormal). The participant reporting radiographers and consultant radiologists had the proforma explained. Particular emphasis was given to the confidence scale used for abnormalities detected in the study. The importance of assigning each abnormality a confidence score, and how to include this in the body of their free text report, was reinforced.

Feedback from the training was given to each participant by the chief investigator, in line with best practice for free-response observer studies.[244]

3.2.7 Obtaining the radiology reference standard diagnosis

3.2.7.1 Image interpretation to obtain the reference standard diagnosis

The pseudoanonymised chest X-rays were independently interpreted by two expert chest consultant radiologists, blinded to the clinical report provided at the time of examination. They were asked to indicate on a proforma if they believed the chest X-ray was normal or abnormal, to provide a supporting free-text report, allocate a disease category for the X-ray and assign a conspicuity (difficulty) rating for the most important abnormality for abnormal cases (see Appendix 15). In order to perform jack-knife alternate free-response receiver operator characteristic (JAFROC) curve analyses the expert chest radiologists who produced the reference standard diagnosis also provided a list containing the location of all abnormalities for the abnormal cases. For each case, each lesion was assigned a number (e.g. Case 1, Lesion 1, Lesion 2 etc.).

The expert chest consultant radiologists had access to all pertinent imaging including previous chest X-rays, follow up X-rays and cross-sectional imaging (e.g. computed tomography) where available. They were provided with the patient demographics (age, gender), referral source (emergency department, out-patient, in-patient) and the clinical history provided by the referring clinician at the time of the request. A list of incidental findings to be considered normal, based on the work of Robinson *et al*. was agreed in advance and was available at time of interpretation (see Appendix 15).[188] The free-response paradigm requires that an acceptance radius be determined prior to the study commencing.[243 249] For this study, the acceptance radius was pre-defined zonal criteria, chosen to reflect the system used in clinical practice.[250] A diagram of the defined zones was included with the proforma (see Appendix 15) and is consistent with that employed in other free response studies which have examined chest X-ray interpretation.[251] The expert chest radiologists were asked to list the location and diagnosis of all abnormalities visible on the chest X-ray. Lesions that were missed at the time of initial interpretation yet visible with retrospective review following follow up or cross-sectional imaging were included in the study. Lesions only visible on subsequent imaging were excluded. This is in line with best practice recommendations for diagnostic accuracy studies.[77 252 253] Image interpretation occurred over five sessions.

Using JAFROC analysis for free-response data, each abnormal image has a mandatory combined weight of 1.0 regardless of the number of lesions.[241] In order to simulate clinical practice as closely as possible, images with more than one abnormality had each lesion weighted for clinical importance. This weighting was the consensus decision of an expert chest consultant radiologist and a professor of respiratory medicine, using a combination of radiology and clinical case note review data.

3.2.7.2 Report comparison – reference standard diagnosis

Two independent arbiters compared the findings of the expert chest consultant radiologists. For each case, the list and location of any (all) abnormalities produced by each expert chest radiologist were compared by two independent arbiters. Arbiters were blinded to the source of the report and did not have access to the images, patient demographics or clinical history. The template is included in Appendix 16. Cases where both arbiters were in agreement that expert chest radiologists identified the same abnormalities and reached an identical diagnosis were included in the image test bank.

For the purposes of the study, the following definition is used:

Arbiter: trained professional who used predefined criteria to compare the observer practitioner report to the reference standard diagnosis

adapted from Brealey and Scally (2008)[63] and Robinson *et al*. (1999)[188]

As required for diagnostic accuracy studies that use the free-response approach, a robust reference standard diagnosis (ground truth) for all cases is required for analysis.[64 241] The diagnosis reached by the expert chest consultant radiologists was taken as the 'truth' regarding the disease state for that patient. A robust reference standard was created by these expert consultant radiologists.

3.2.8 Interpretation of the chest X-ray image bank to establish diagnostic accuracy of participants

3.2.8.1 Image interpretation

The bank (n=106) of chest X-rays compiled by the expert chest consultant chest radiologists was given to the participant reporting radiographers and consultant radiologists. Each participant provided independent interpretations for each case, blinded to both the clinical report and the reference standard diagnosis. The participants (consultant radiologists and reporting radiographers) had access to previous chest X-rays but not any other imaging investigation. They were provided with the patient demographics (age, gender), referral source (emergency department, out-patient, in-patient) and the clinical history provided by the referring clinician at the time of the request. A list of incidental findings to be considered normal was available at time of interpretation. Localisation of any abnormalities was performed using the same acceptance radius used when determining the reference standard diagnosis; pre-defined zonal criteria.[250] A diagram was included with the proforma (see Appendix 17).

Participants were asked to indicate whether they interpret the image as normal or abnormal. A free report was also required, and described the salient features, a diagnosis or differential diagnosis (if appropriate) and any specific recommendations regarding further patient management. To enable JAFROC analysis, participants were asked to assign a confidence score (1 – 4; 1=uncertain, 2=possibly abnormal, 3=probably abnormal, 4=definitely abnormal) for each abnormality identified on the image.[241 243 244 246] Convention in free-response diagnostic accuracy studies is that normal cases are assigned zero confidence score by default,[64 241] and was the method used in the current study. Participants were advised to include these confidence scores in parentheses within the free text report in order to minimise disruption to the normal clinical reporting task. Any image deemed normal by the participants was assigned a confidence score of zero by the investigator; this did not need to be indicated by the participant.

Image interpretation occurred over two sessions, with half (53 cases) included in each session. In order to minimise participant inconvenience and to mimic normal clinical practice as closely as possible, these occurred on the same day, separated by a one hour break. Participants were randomised for image bank and ascending/descending sequence.[240] The elevated disease prevalence used in image test banks may alter observer performance, often due to increased sensitivity, when compared to clinical practice.[254] Although the normal:abnormal ratio of 1:1 is higher than would be expected in clinical practice, recent research has shown that higher disease prevalence does not alter observer performance when interpreting chest X-rays in a structured environment.[255] The disease prevalence of the image bank used in the current study

(50% abnormal) is not dissimilar to the disease prevalence found at audit (40% abnormal) at the clinical centre from which the cases were drawn (Appendix 3).

3.2.8.2 Report comparison – participant observers for diagnostic accuracy study

Two independent arbiters compared the free-text reports produced by the participant observers and identified each mark-rating pair (a localised abnormality assigned a confidence score).[241 243 246] The mark-rating pair was compared against the reference standard diagnosis for each case using a proforma (Appendix 18). Abnormal findings were identified for each image. Abnormalities which corresponded to the location and diagnosis (lesion) confirmed by the reference standard were designated a lesion localisation (LL) event.[241] All abnormalities that were either (i) within the acceptance radius but reach an incorrect diagnosis or (ii) outside the acceptance radius were designated a non-lesion localisation (NLL) event.[241] The confidence score for each lesion localisation and non-lesion localisation was recorded. Training was given by the chief investigator to enable accurate comparison by the independent arbiters. This consisted of 5 practice cases, presented in conjunction with a short review of the pertinent literature[241 243 244 246 256] and the guidance produced by the JAFROC software package (http://www.devchakraborty.com/index.php).[257]

3.2.9 Data analysis – diagnostic accuracy study

3.2.9.1 Data collation – transfer of reference standard diagnosis to participant image bank

Cases in which both expert consultant chest radiologists were deemed to be in agreement by the independent arbiters were assigned a unique case reference number and had the abnormality list (location and diagnosis) extracted and coded by the chief investigator. This was entered into a Microsoft Excel 2010 spreadsheet (see Appendix 19) and formed the 'truth' for that case. Cases with more than one abnormality had a number assigned to each lesion. The chief investigator allocated lesion weighting, agreed for each case by an expert chest radiologist and a professor of medicine.

The reports produced when obtaining the reference standard diagnosis were compared to the file report by the chief investigator. Abnormalities identified by the expert chest radiologists during the study but not at the time of clinical interpretation were identified. The research team (chief investigator, consultant chest radiologist and professor of medicine) reviewed these cases. Consensus was reached regarding the final diagnosis using all available clinical and radiological information. There were no clinically significant confirmed abnormalities missed in the clinical report and subsequently identified which had not already been identified during routine care. There were no cases which required notification of the named consultant for the patient or that required notification of the patient according to Trust procedures[258 259] or under the Duty of Candour.[260]

3.2.9.2 Diagnostic accuracy of participant observers

Each abnormality (both lesion localisations and non-lesion localisations) identified from the free text reports by the independent arbiters for each case was extracted along with the assigned confidence score. This was coded and entered into a Microsoft Excel 2010 spreadsheet by the chief investigator (see Appendix 20).

3.2.10 Statistical analysis – diagnostic accuracy study

3.2.10.1 Statistical analysis – reference standard diagnosis

Inter-observer agreement between the two expert chest radiologists was determined using the Kappa statistic (K) to determine consistency among observers and 95% confidence intervals constructed using SPSS (IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY). The null hypothesis (standard superiority) was rejected if the p value produced was less than 0.05. The level of inter-observer agreement was quantified according to the criteria established by Landis and Koch,[261] where:

K = 0.01 – 0.20	slight agreement
K = 0.21 – 0.40	fair agreement
K = 0.41 – 0.60	moderate agreement
K =0.61 – 0.80	substantial agreement
K = 0.81 – 1.0	almost perfect agreement

3.2.10.2 Statistical analysis - participant observers

A free-response paradigm was used to determine the diagnostic accuracy of the participant observers (consultant radiologists and reporting radiographers). Each abnormality identified was assigned a lesion localisation or non-lesion localisation score according to the criteria outlined above (Chapter 3.2.8.2). The JAFROC methodology was again used (detailed in Appendix 12.4.1.2) to determine the performance of each observer and each observer professional group using JAFROC software (version 4.2, http://www.devchakraborty.com/index.php).[257]

A figure of merit (FoM), the summary performance score, was calculated for each individual practitioner and an average for each professional group. There is no established benchmark for acceptable performance for chest X-ray interpretation. The performance of reporting radiographers is expected to be comparable to consultant radiologists in the interest of optimal care and patient safety.[44] As outlined in Appendix 12.6 a non-inferiority approach was used in the current study. To facilitate this, the average performance of the cohort of consultant radiologists was used as the baseline.

It is important to differentiate between clinical and statistical significance in healthcare research.[129] Clinical significance is the level of difference between observers that would be of no clinical consequence in everyday practice has been derived from the literature. The CONSORT extension for non-inferiority studies explicitly states that the margin of non-inferiority should be decided *a priori*; to perform accurate sample size calculations and to allow robust statistical analysis.[48] Review of the literature identifies significant variability in chest X-ray interpretation accuracy.[65-67 70 71 74 76 77 82 132 174 188 194 196 197 204 205 207 214 226 240 262-271] Based on consensus from the literature, guidance from relevant professional bodies and the clinical evaluation, an inter-observer variability of less than 10% in chest X-ray interpretation would be accepted in clinical practice. A 10% difference in diagnostic accuracy was used as the pre-defined margin of non-inferiority in the study.

The free-response pseudovalues produced by JAFROC analysis for each participant were calculated. Utilising the non-inferiority approach, the null hypothesis that the FoM for the consultant radiologists exceeds the pre-defined clinically significance level of 10% was rejected if the p value was less than 0.05 for a one-tailed test.[238] Ninety-five percent confidence intervals were constructed for the FoM for each professional group (consultant radiologists and reporting radiographers).

To facilitate comparison with previous studies that have used receiver operator characteristic (ROC) curves, JAFROC (version 4.2)[257] was used to produce inferred ROC curves. The sample size was sufficient to perform standard ROC analysis and to detect a statistical difference in performance between consultant radiologists and reporting radiographers, if one exists.[242] The area under the curve (AUC) for consultant radiologists and reporting radiographers was calculated and compared. Utilising the non-inferiority approach, the null hypothesis that the AUC for the consultant radiologists exceeded the pre-defined clinical significance level of 10% was rejected if the p value was less than 0.05 using a one-tailed test.[238]

3.3 Part 2 – Influence on clinicians' diagnostic decision-making

Primary research question and hypothesis

Primary Research Question Is there a clinically significant difference between the influence that consultant radiologist and reporting radiographer chest X-ray reports have on clinicians' diagnostic decision-making?

Alternative Hypothesis [non-inferiority approach (Appendix 12.6)]: That there is no clinically significant difference in the influence that chest X-ray reports produced by consultant radiologists or reporting radiographers have on clinicians' diagnostic decision-making

Null Hypothesis [non-inferiority approach]: That consultant radiologist chest X-ray reports will have a clinically significant difference on clinicians' diagnostic decision-making when compared to reporting radiographer's chest X-ray reports.

Secondary research question and hypothesis

Secondary research question Is there a difference between the influence that chest X-ray reports have on clinicians' diagnostic decision-making due to clinician experience?

Alternative Hypothesis [standard superiority] That clinician experience will moderate the influence that chest X-ray reports (of both consultant radiologists and reporting radiographers) have on diagnostic decision-making

Null Hypothesis [standard superiority] That clinician experience will not moderate the influence that chest X-ray reports (of both consultant radiologists and reporting radiographers) have on diagnostic decision-making

3.3.1 Methodology – influence on diagnostic decision-making

Part 2 is a retrospective quasi-experimental study that compared the impact of reporting radiographer and consultant radiologist chest X-ray reports on clinicians' decisions, using a non-inferiority approach. This examined level 3 efficacy as outlined by Fryback and Thornbury.[57] The methods used in part two of the study are outlined in Figure 3.2.

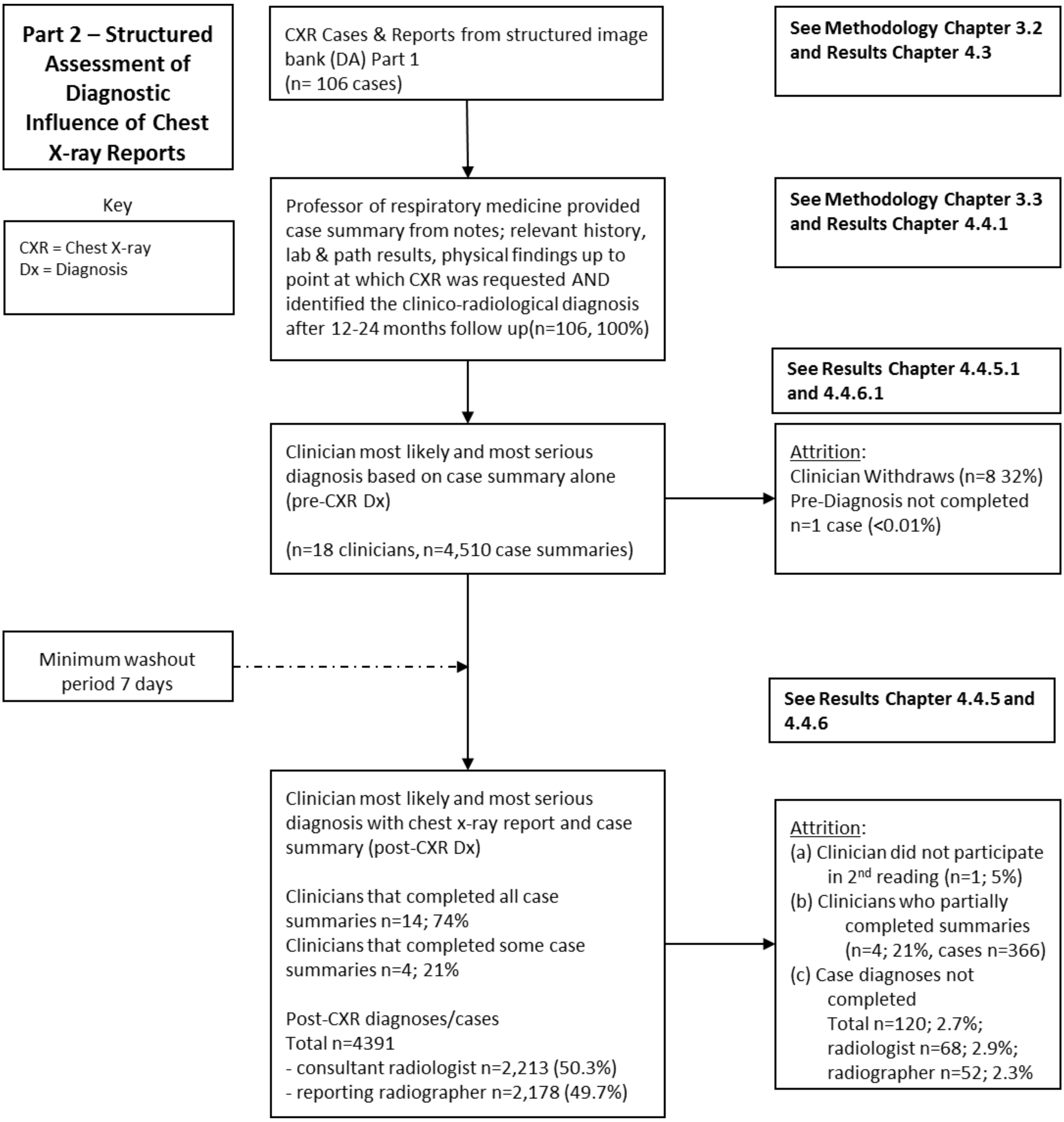


Figure 3.2 Study design flow diagram for Part 2: Assessment of Diagnostic influence of chest X-ray reports

3.3.2 Case selection and allocation

The cases, taken from a consecutive series from a single acute London district general hospital, and in which the reference standard agreed the diagnosis in the Part 1 study were used in the Part 2 analysis. The chest X-ray reports produced by the reporting practitioners (consultant radiologists and reporting radiographers) in Part 1 were used in the analyses.

The case notes for the above cases were retrieved from Medical Records, with case notes available for all (n=106) patients. A professor of medicine/consultant respiratory physician performed a comprehensive review of all information. A final diagnosis was determined and all relevant clinical information was extracted. A short case summary was produced for each case.

Each case was randomly assigned, using a random number generator (Microsoft Excel), and the brief case summary reviewed by at least 2 clinicians without the chest X-ray report. After a washout period (minimum 7 days), the case summaries and corresponding chest X-ray report were reviewed by a clinician. Clinician allocation was quasi-randomised to ensure that clinicians only reviewed cases for which they reviewed the pre-chest X-ray case summary.

Case summaries that did not have both the pre and post chest X-ray report data completed were excluded from analysis.

3.3.3 Sample size calculation – influence on diagnostic decision-making study

Based on the work of Lusted, the only study to examine the influence of chest X-ray reports on clinicians' diagnostic thinking,[233] and in consultation with a statistician, it was estimated that the consultant radiologist chest X-ray reports would produce a new (not mentioned as a pre-chest X-ray) diagnosis in 20% of cases. Utilising a pre-defined clinically insignificant difference of less than 10%, a non-inferiority study required 198 cases to produce an 80% power of sample with a 5% possibility of a Type I error, according to the formula outlined by Scally and Brealey (p.244).[272]

$$n = \frac{(1.645 + 0.842)^2[0.2 \times 0.8 + 0.2 \times 0.8]}{[0 - 0.10]^2} = 198$$

To provide an estimate of the required sample size for different proportions of consultant radiologist chest X-ray reports producing a new diagnosis the above calculation has been performed using 5% (n=58 cases) and 10% (n=111 cases) of reports producing a new diagnosis with the same margins of non-inferiority. To ensure the study was appropriately powered the conservative sample size (20% new diagnosis, n=198 cases) was used.

The study was effectively clustered by each reporting practitioner (consultant radiologist or reporting radiographer) who produced multiple reports.[273 274] This lack of independence was accounted for with a design effect, a method outlined by several authors.[273 274]

$$\text{design effect} = 1 + (m - 1) \times \text{ICC}$$

where $m = number\ of\ observers$ and $ICC = intracluster\ correlation\ coefficient$

A revised sample size of 970 reports was required for each professional group (consultant radiologists and reporting radiographers). A total of 2,213 chest X-ray reports were generated in Part 1 (diagnostic accuracy) of the study; 1,055 consultant radiologist and 1,158 reporting radiographer. Thus, an adequate sample for clinician review in Part 2 (diagnostic influence) was produced.

3.3.4 Study participants – influence on diagnostic decision-making study

Inclusion criteria

- Current medical staff, holding GMC registration and employed at the Trust

Exclusion criteria

- Medical staff employed in a surgical post

All participants were invited to participate and were free to withdraw at any time via direct approach at clinico-radiological team meetings and email of relevant medical teams at the Trust. Participant information sheets, appropriate to the clinician taking part in the study were provided to each participant (see Appendix 9). The chief investigator obtained signed informed consent prior to participant enrolment (see Appendix 11). Each participant was assigned a unique study identifier to ensure anonymity. A small honorarium was paid in acknowledgement of the time and effort required to participate (chapter 3.1).

Participant clinicians' were recruited, using purposive sampling, from the Trust, the same hospital from which the CXR cases were selected. All clinicians within the Trust were approached to participate, and participants were selected according to order of response, until the required number was met. Clinicians were recruited through staff email (Appendix 21) and multidisciplinary team meetings.

For the purposes of this study, the consultant grade comprised of staff that held a substantive consultant post within the hospital and specialist registration (general or emergency medicine) with the General Medical Council (GMC).

Specialist registrars were qualified medical practitioners (GMC registered) on a registered training programme, while junior medical staff were qualified medical practitioners (GMC registered) who had not yet begun specialist training. A professor of medicine and consultant physician in respiratory medicine, employed by the Trust and who held specialist registration (general and respiratory medicine) with the General Medical Council, was recruited to produce the succinct case summary from the patient notes.

General demographic details were recorded for each participant. These consisted of:

- Area of specialism
- Experience (Consultant, Specialist Registrar, Junior)

3.3.5 Establishing clinicians' diagnoses

A professor of medicine and consultant respiratory physician produced a succinct summary for cases used in Part 1. The summary was based on the clinical information available up until the point that the chest X-ray used in Part 1 was requested. The consultant physician also identified the final diagnosis for that patient from the medical records, which was based on all available clinical, histological and radiological information.

Pre and post-CXR report proformas, based on work previously conducted,[136 162-164 172 233 275-277] have been designed, and were piloted to ensure reliability and validity prior to commencement of the study (Appendix 22 and 23). They contained patient demographics, referral source and the case summary. Attempts were made to recruit eight clinicians for each level of experience (consultant, registrar, junior medical staff), from a range of specialities. In total 27 clinicians were recruited and consented to participate, with eight withdrawing prior to data collection (30%) and one only completed the pre-chest X-ray data. Eighteen clinicians were randomly assigned to independently review the case summaries.

The outcome measures, namely a change in diagnosis or diagnostic confidence, utilised in this study have been derived from previously validated measures.[162-164 233 277] Clinicians were asked to select independently the most likely and the most serious diagnosis for each case. The most likely diagnosis is self-explanatory; the most serious diagnosis was defined as the "*condition that the clinician would not want to miss in this patient, even if it is very unlikely*"(p.21).[233] The list of diagnoses available for the most likely and most important are identical, and were compiled based on an audit of most frequent discharge diagnoses in which a chest X-ray was performed at the host clinical site. This method of 'pruning' has been demonstrated not to adversely bias results.[146 163]

A continuous measurement scale (0 -100; 0=very unlikely, 100=certain) was used to measure the clinicians' confidence in their diagnostic decisions. Confidence was measured prior to and in conjunction with a chest X-ray report.[233 277] At least one clinician assessed each chest X-ray report. Post-chest X-ray cases were only given to clinicians who had reviewed the initial case summary. This occurred over a minimum of two sessions.

3.3.6 Data analysis – influence on diagnostic decision-making study

3.3.6.1 Data collation

Cases that produced a new post chest X-ray report diagnosis or an alteration in the confidence of an existing decision were identified, for both most likely and most serious diagnoses. The most likely post chest X-ray report diagnosis was compared to the final diagnosis for accuracy, with confidence corrected according to the Tsushima method.[277] Data were entered into a spreadsheet prior to analysis (Appendix 24).

3.3.7 Statistical analysis – influence on diagnostic decision-making study

This study was conducted in a controlled setting, using images obtained for clinical reasons. The decisions of the clinicians did not have any influence on patent care or outcomes. The proportion of reporting radiographer chest X-ray reports that produced a positive influence on clinicians' diagnostic decision-making were compared to the proportion of consultant radiologist reports. A pre-defined non-inferiority margin of 10% was utilised. Statistical analysis was performed using SPSS (IBM, version 21). The null hypothesis rejected if the p value produced was less than 0.05. Bonferroni's correction for multiple observers was not required, as the lack of independence of data (clustering by reporting practitioner) had been accounted for in the sample size calculations and randomisation of cases to reviewing clinician.

(i) Pre- and post-chest X-ray report diagnostic confidence (uncorrected)

A one way t-test was used to compare the clinician pre- and post-chest X-ray uncorrected diagnostic confidence (continuous data) both prior to and in conjunction with a consultant radiologist or reporting radiographer chest X-ray report. The pre-defined non-inferiority margin of 10% of the average consultant radiologist diagnostic confidence was utilised.[144 277]

(ii) Pre- and post chest X-ray report diagnostic confidence (corrected)

The post chest X-ray clinician diagnosis was compared to the final diagnosis and corrected for accuracy using the methods outlined by Tsushima and colleagues.[277] A one-way t-test was used to compare the clinician pre- and post-chest X-ray corrected diagnostic confidence (continuous data) both prior to and in conjunction with a consultant radiologist or reporting radiographer chest X-ray report. The pre-defined non-inferiority margin of 10% of the average consultant radiologist diagnostic confidence was utilised.[144 277]

(iii) Reports producing a new diagnosis

Analysis of the proportion of reporting radiographer and consultant radiologist reports that resulted in a new diagnosis (most likely or most serious) was conducted using a Chi-square test.[144 163 164 278]

(vii) Difference in proportion of new diagnoses between clinician of different experience

A Chi-squared test was used to examine if there was a difference in the proportion of reporting radiographer and consultant radiologist reports which produced a change in diagnosis between clinicians' of different experience (consultant, registrar, junior medical staff), using standard hypothesis testing (Null hypothesis = no difference between clinician grade).[144 163 172 278]

Chapter 4 – Results

4.0 Introduction of results

The study carried out consisted of two parts. Part 1 was an investigation that compared the diagnostic accuracy of radiologist and radiographer chest X-ray reports. Part 2 compared the influence of chest X-ray reports on the diagnostic decision-making of referring clinicians. The results of the two investigations and the inter-relationship of the findings are reported in this chapter.

4.0.1 Summary of the purpose of Part 1 of the study

An essential component of Part 1 of this study was the need to create a rigorous image bank of chest X-rays with robust reference standard diagnoses. Section 4.1 details the construction of the image bank and investigated the agreement between the reports of the expert chest consultant radiologists to establish the reference standard diagnosis for cases included in the image bank. The expert chest radiologist reports were independently compared by two arbiters for agreement, to test the hypothesis that there was good agreement between these experts.

> Research Question: What is the level of agreement between the expert consultant chest radiologists used to establish the reference standard diagnosis?
>
> Alternative Hypothesis [superiority approach]: There will be good agreement between the expert chest consultant radiologists when interpreting chest X-rays.
>
> Null Hypothesis [superiority approach]: There will be poor agreement between the expert chest consultant radiologists when interpreting chest X-rays.

All chest X-rays included in the study were retrieved from a randomised, stratified sample from clinical practice as described in Section 3.2.3. The clinical reports provided for each X-ray at the time of the original examination, by either a consultant radiologist or a reporting radiographer, were compared to the expert reports for agreement. The results of this are reported in Section 4.2.

Section 4.2 presents the results of the examination of the hypothesis that there would be similar agreement between the original clinical reports and the expert chest radiologist reports.

> Research Question: What is the agreement of expert chest radiologists with chest X-ray reports produced in clinical practice by consultant radiologists and reporting radiographers?
>
> Alternate Hypothesis [superiority approach]: That there will be comparable agreement between expert chest radiologists and the chest X-ray reports produced in clinical practice by consultant radiologists and reporting radiographers
>
> Null Hypothesis [superiority approach]: That agreement between expert chest radiologists will be greater for the chest X-ray reports produced in clinical practice by the consultant radiologists when compared to the chest X-Ray reports of the reporting radiographers

Section 4.3 reports the results of the primary purpose of Part 1 of this study which was to compare the diagnostic accuracy of two groups, a cohort of reporting radiographers (RR) and a cohort of consultant radiologists (CR) when interpreting the reference standard bank of adult chest X-rays (n=106). As explained in Appendix 12.6 the non-inferiority approach was used.

Research Question: Is the accuracy with which a group of consultant radiologists interprets a bank of adult chest X-rays comparable to the accuracy of a group of reporting radiographers?

Alternative Hypothesis [non-inferiority approach]: That there is no clinically significant difference in the accuracy of adult chest X-ray interpretation between consultant radiologists and reporting radiographers.

Null Hypothesis [non-inferiority approach]: That the accuracy of adult chest X-ray interpretation of consultant radiologists will be significantly (both clinically and statistically) superior when compared to the accuracy of reporting radiographers.

4.0.2 Summary of the purpose of Part 2 of the study

Section 4.4 reports the results of this examination in which the reporting radiographer and consultant radiologist chest X-ray reports were compared for changes in diagnosis and diagnostic confidence of referring clinicians. The non-inferiority approach was again used. The reporting radiographer and consultant radiologist chest X-ray reports, produced in Part 1, were compared for changes in diagnosis and diagnostic confidence of referring clinicians, using a non-inferiority approach to examine the hypothesis that there was no clinically significant difference.

Research Question: Is there a significant difference between the influence that reporting radiographer and consultant radiologist chest X-ray reports have on clinicians' diagnostic decision-making?

Alternative Hypothesis [non-inferiority approach]: That there is no clinically significant difference in the influence that reporting radiographer chest X-ray reports have on clinicians' diagnostic decision-making when compared to consultant radiologist reports.

Null Hypothesis [non-inferiority approach]: That consultant radiologist's chest X-ray reports will have a clinically important difference on clinicians' diagnostic decision-making when compared to reporting radiographer's chest X-ray reports.

Results of Part 1 of the study

4.1 Expert radiologist agreement in construction of a robust image test bank

A random stratified sample of adult chest X-rays, previously performed for clinical reasons at a single London acute district general hospital over a consecutive twelve month period, were interpreted independently by two consultant radiologists with subspecialist interest in thoracic imaging (consultant chest radiologists; CC1/CC2). A total of 193 cases were included as described in Chapter 3 (Section 3.2.2 'Inclusion criteria'). Each expert radiologist interpreted 187 (97%; CC1) and 186 (96%; CC2) examinations respectively. The small number of cases not interpreted by each expert (CC1 6 cases; CC2 7 cases) occurred due to expert oversight. There was no pattern to the cases only interpreted by a single expert. Thirty-two cases (32/193, 17%) were selected from radiology discrepancy meetings previously held at the study site. Steps were taken to ensure the image bank replicated the typical case mix likely to be found in clinical practice as closely as possible, using local diagnosis audit to compare with population data available for England (see Chapter 3.2.3).

4.1.1 Referral sources for cases included in the study

The referral sources for the cases included in the study are shown below in Fig 4.1. As can be seen, the majority were from referrals from the emergency department, with inpatients and outpatient clinic patients' together accounting for just under half of the cases.

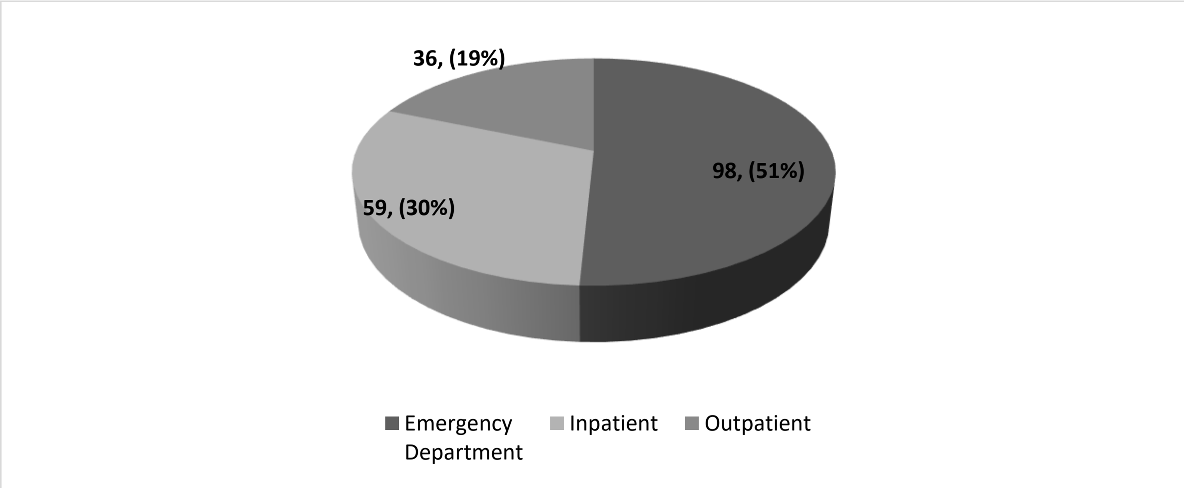


Figure 4.1 Source of referral associated with the cases.

Table 4.1 shows the prior and follow-up imaging relative to the cases that were used to develop the reference standard image bank. For this study, expert radiologists had access to all imaging available including chest X-rays and cross-sectional imaging (Chapter 3.2.7.1). For most cases, there was previous or follow-up imaging, with some having both. A small number of cases (n=32; 17%) had no other imaging either before or after the chest X-ray included in the bank.

	Number of Cases
Previous CXR	38 (20%)
Follow Up CXR	34 (18%)
Both Previous and Follow Up CXR	89 (46%)
Neither Previous nor Follow Up CXR	32 (17%)

Table 4.1 Proportion of cases with previous and/or follow up imaging

4.1.2 Normal – Abnormal agreement of Expert Radiologists

Of the original cases drawn from clinical practice and teaching files, 180 were interpreted by both expert chest radiologists. As outlined in Chapter 3.2.7.1, expert radiologists were required to make a normal:abnormal decision for all cases. Their agreement figures for the dichotomous normal abnormal decisions are presented in Table 4.2.

	Agree Normal (CC1-CC2)	Agree Abnormal (CC1-CC2)	Total Cases	Kappa (95% CI)
Referral Source				
ED	28 (30%)	35 (38%)	92	0.42 (0.28 – 0.57)*
IP	9 (17%)	32 (62%)	52	0.49 (0.25 – 0.73)*
OP	15 (42%)	12 (33%)	36	0.52 (0.26 – 0.63)**
OVERALL	52 (29%)	79 (44%)	180	0.48 (0.36 – 0.59)*

Table 4.2 Normal-Abnormal agreement between expert radiologists for chest X-rays from different referral sources
ED = Emergency Department, IP = Inpatient, OP = Outpatient *p<0.0001; **p=0.001

The expert radiologists (CC1 and CC2) agreed on 52 (on 180; 29%) normal and 79 (of 180; 44%) abnormal X-rays, with moderate agreement. The source of referral of the chest X-rays did not appear to influence results.

4.1.3 Influence of additional imaging availability on agreement between expert radiologists

As described in Chapter 3.2.7 previous and or follow-up imaging was available to the expert chest radiologists. This did not appear to influence their moderate agreement, as shown in Table 4.3.

	Agree Normal (CC1-CC2)	Agree Abnormal (CC1-CC2)	Total Cases	Kappa (95% CI)
With Previous Imaging	11 (6%)	16 (9%)	34 (19%)	0.60 (0.35 – 0.85)*
With Follow Up Imaging	9 (5%)	14 (8%)	31 (17%)	0.50 (0.24 – 0.77)**
Both Previous and Follow Up	15 (8%)	45 (25%)	83 (46%)	0.41(0.23 – 0.58)*
Neither	17 (9%)	4 (2%)	32 (18%)	0.25 (0 – 0.53)***
OVERALL	52 (29%)	79 (44%)	180	0.48 (0.36 – 0.59)*

Table 4.3 Normal-Abnormal agreement between expert radiologists for chest X-rays with previous and follow up imaging
*=p<0.0001; **=p=0.001; ***p=0.075

4.1.4 Influence of diagnosis on normal-abnormal agreement between expert radiologists

The experts were required to assign a disease category for each abnormal case: infection, cardiac/pulmonary oedema, malignancy and 'other' as detailed in Chapter 3.2.7. The agreement between expert radiologists when assigning the disease category are presented in Table 4.4.

		Expert 2 (CC2) Diagnosis					
		Normal	Infection	Cardiac	Malignant	Other	
Expert 1 (CC1) Diagnosis	Normal	**52**	1	1	0	0	
	Infection	23	**14**	8	5	2	
	Cardiac	12	0	**22**	0	1	
	Malignant	1	0	1	**7**	0	
	Other	11	2	0	1	**16**	
Total Agreement		**52**	**14**	**22**	**7**	**16**	111 (of 180)

Table 4.4 Normal-abnormal agreement between expert radiologists for assigned disease category

As may be shown in Table 4.4 there was agreement on disease category assignment in 111 (of 180; 62%) of the 180 cases and disagreement in 69 (of 180; 38%). Of the cases which produced disagreement in disease category between the experts, the largest source of discrepancy between the expert radiologists were cases interpreted as normal by expert 2 yet abnormal by expert 1 (47 cases, 26%). Infection was also a source of discrepancy between the experts; 8 cases (4%) compatible with infection by expert 1 were diagnosed as cardiac disease by expert 2 and 5 cases (3%) where expert 1 diagnosed infection but malignant by expert 2.

In addition to categorising each case as normal or abnormal, and assigning a disease category, the expert chest radiologists were required to identify and localise each abnormality for abnormal cases. Cases were deemed to be in complete concordance only when two independent arbiters agreed that all abnormalities were identified and localised by both expert chest radiologists (Chapter 3.2.7.2). This detailed information regarding the locality of each abnormality on included abnormal cases also facilitated alternate free response data collection for the diagnostic accuracy study.

4.1.5 Agreement between arbiters when assessing expert chest radiologist reports for agreement of all findings (complete report concordance)

The agreement of each arbiter for cases rated normal or abnormal by each expert was determined. As detailed in Chapter 3.2.7, cases that were interpreted by the expert chest radiologists as normal did not require them to provide further information. Both arbiters agreed that the abnormalities described by expert 1 in three instances were included on the list of insignificant findings according to study protocol (Chapter 3.2.7), and thus should have been considered normal (calcified granuloma, thoracic scoliosis, previous surgery). A single case rated as normal by expert 1 (upper lobe fibrosis from previous tuberculosis) was to be considered abnormal for the purposes of this study. The normal-abnormal decisions of expert

1 for these for cases were corrected according to the study protocol and included in the image bank. Agreement between the arbiters when assessing the expert radiologist reports for concordance is presented in Table 4.5.

		Arbiter 2		Total
		Disagree	Agree	
Arbiter 1	Disagree	55 (31%)	6 (3%)	61
	Agree	13 (7%)	106 (59%)	119
Total		68	112	180

Table 4.5 Agreement between arbiters when assessing expert radiologist report concordance

In 59% (106 of 180) of cases both arbiters deemed both expert reports to be in complete concordance. Both arbiters found discordant reports in 31% (55 of 180) of cases. There was disagreement between arbiters in 19 cases (10%). Substantial agreement was found between the arbiters, Kappa statistic (K)=0.77 (95%CI 0.67 – 0.87; $p<0.0001$). The source of referral of the chest X-ray (Chi Square χ=2.265, p=0.322) and the availability of additional imaging (previous χ=01.534, p=0.215; follow up χ=0.450, p=0.502) did not appear to influence agreement.

4.2 Agreement between the expert chest consultant radiologist reports and the reports provided by reporting radiographers and consultant radiologists in clinical practice

The chest X-rays included in the study were taken from a retrospective series performed for clinical reasons as part of patient management and as such, had clinical reports provided at the time of examination. In the department where case selection occurred, there were both consultant radiologists and reporting radiographers providing clinical reports for adult chest X-rays. The findings of the expert chest consultant radiologists were compared to the clinical reports by two independent arbiters, blinded to reporting source (Chapter 3.2.7).

4.2.1 Normal – Abnormal agreement between the expert radiologists and clinical reports

Of the original 193 cases drawn from clinical (n=161, 78%) and discrepancy meeting files (n=32, 17%), 180 were reported by both CC1 & CC2. The 12 cases not reported by both expert chest radiologists were excluded from further analysis. Of the 180 reports included, 52% (93 of 180) were produced by consultant radiologists and 48% (87 of 180) by reporting radiographers. For the dichotomous normal/abnormal decision, the agreement figures between expert radiologists and the clinical reports of consultant radiologists and reporting radiographers are presented in Table 4.6. Agreement was categorised according to Landis and Koch.[261]

	Agree Normal (Expert-Clinical)	Agree Abnormal (Expert-Clinical)	Total Cases	Kappa (95% CI)
Reporting Radiographer				
Expert CC1	25	45	87	0.59 (0.42 – 0.76)*
Expert CC2	35	35	87	0.62 (0.43 – 0.78)*
Consultant Radiologist				
Expert CC1	23	53	93	0.60 (0.44 – 0.77)*
Expert CC2	35	40	93	0.62 (0.46 – 0.77)*
OVERALL				
Expert CC1	48	98	180	0.60 (0.48 – 0.72)*
Expert CC2	70	75	180	0.62 (0.51 – 0.73)*

Table 4.6 Normal-abnormal agreement between expert radiologists and the clinical report
*=p<0.0001

Agreement for clinical reports (Table 4.6) is as follows; each expert radiologist agreed on 48 (of 180, 27%; CC1) and 70 (of 180, 39%; CC2) normal clinical reports; and 98 (54%; CC1) and 75 (42%; CC2) abnormal clinical reports, with moderate to substantial agreement.[261] The reporting practitioner (consultant radiologist/reporting radiographer) providing the clinical report did not appear to influence results. All agreement rates were categorised as substantial (K>0.6).

4.2.2 Influence of additional imaging availability on agreement between expert radiologists and the clinical report

As the X-rays were selected from a retrospective series, some patients had previous or follow up imaging, which was made available to the expert radiologists. A summary of the included cases are presented in Table 4.7.

	Agree Normal (Expert-Clinical)	Agree Abnormal (Expert-Clinical)	Total Cases	Kappa (95% CI)	p value
With Previous Imaging					
Reporting Radiographer					
Expert CC1	9	6	21	0.46 (0.15 – 0.78)	0.012*
Expert CC2	13	6	21	0.79 (0.51 – 1)	<0.0001*
Consultant Radiologist					
Expert CC1	3	10	14	0.81 (0.45 – 1)	0.001*
Expert CC2	4	8	14	0.70 (0.32 – 1)	0.006*
With Follow Up Imaging					
Reporting Radiographer					
Expert CC1	2	9	12	0.75 (0.29 – 1)	0.007*
Expert CC2	2	7	12	0.44 (0 – 0.9)	0.067
Consultant Radiologist					
Expert CC1	4	7	18	0.23 (0 – 0.64)	0.196
Expert CC2	8	5	18	0.43 (0 – 0.86)	0.066
Both Previous and Follow Up Imaging					
Reporting Radiographer					
Expert CC1	7	27	40	0.60 (0.31 – 0.89)	<0.0001*
Expert CC2	9	21	40	0.47 (0.21 – 0.73)	0.001*
Consultant Radiologist					
Expert CC1	5	31	43	0.50 (0.18 – 0.81)	<0.0001*
Expert CC2	11	24	43	0.61 (0.37 – 0.84)	<0.0001*
Neither					
Reporting Radiographer					
Expert CC1	7	3	14	0.43 (0 – 0.83)	0.051
Expert CC2	11	1	14	0.44 (0 – 1)	0.047*
Consultant Radiologist					
Expert CC1	11	5	18	0.75 (0.44 – 1)	0.001*
Expert CC2	12	3	18	0.56 (0.11 – 1)	0.017*

Table 4.7 Influence of additional imaging on normal-abnormal agreement between expert radiologists and the clinical report
*=statistically significant result ($p<0.05$)

Using the criteria of Landis and Koch [261] outlined in Chapter 3.2.10.1, moderate or substantial agreement between expert radiologists and the clinical reports was again found. The exceptions were expert 1-consultant radiologist reports with previous imaging (almost perfect Kappa 0.81) and expert 1-consultant radiologist reports with only follow up imaging (poor, Kappa 0.23). Access to additional imaging did not appear to influence results. There did not appear to be a difference in expert agreement according to whether the report was from a reporting radiographer or a consultant radiologist (Table 4.7).

4.2.3 Complete report concordance between the expert chest radiologists and the clinical reports

Two independent arbiters compared the clinical report to the report of each expert radiologist for concordance. Using the same methodology (see Chapter 3.2.7 and Chapter 4 Section 4.1.5), for the purpose of this study, reports were defined as being in concordance only when both independent arbiters agreed that all abnormalities were identified and localised by both the expert chest radiologists and the clinical reports. Complete concordance was determined between expert radiologists (CC1 and CC2) and between each expert and the clinical reports provided by consultant radiologists and reporting radiographers. The results are presented in Table 4.8.

	Disagree	Agree	Total Cases
Reporting Radiographer			
Expert CC1	38 (44%)	49 (56%)	87
Expert CC2	32 (37%)	55 (63%)	87
Consultant Radiologist			
Expert CC1	39 (42%)	54 (58%)	93
Expert CC2	35 (38%)	58 (62%)	93

Table 4.8 Expert chest radiologist report concordance with the clinical report

Report concordance between each expert radiologist and the clinical report (provided by the consultant radiologists or reporting radiographers) were comparable, with no statistically significant difference found between the proportions (CC1 Chi square χ=0.056, p=0.813; CC2 χ=0.014, p=0.906).

Concordance between the expert radiologists and the clinical reports was stratified by inter-expert agreement, to account for variability between the expert radiologists. The results are presented in Table 4.9.

	Inter-Expert (CC1-CC2)		
	Disagree	Agree	Total Cases
Reporting Radiographer			
Expert CC1	28 (32%)	42 (48%)	87
Expert CC2	21 (24%)	41 (47%)	87
Consultant Radiologist			
Expert CC1	27 (29%)	39 (42%)	93
Expert CC2	25 (27%)	41 (44%)	93

Table 4.9 Complete report concordance between expert radiologists and the clinical report

The reports of expert radiologist CC1 were in concordance with both the other expert radiologist (CC2) and the clinical report for 39 (of 93, 42%) consultant radiologist and 42 (of 87, 48%) reporting radiographer reports. A similar proportion of expert radiologist CC2 reports produced concordance between the expert radiologists and the consultant radiologist (41 of 93, 44%) and reporting radiographer (41 of 87, 47%) clinical reports. When variability between expert radiologists was accounted for, there did not appear to be a significant difference in report concordance between consultant radiologist (McNemar, p=0.701) and reporting radiographer (McNemar, p=0.629) clinical reports.

4.3 Diagnostic accuracy of a cohort of consultant radiologists and reporting radiographers when interpreting a bank of adult chest X-rays

The primary analysis used in this study to determine the diagnostic accuracy of consultant radiologist and reporting radiographer chest X-ray reporting was weighted jack-knife alternate free response receiver operator characteristic curves (JAFROC). This method most closely replicates clinical practice; it allows for multiple abnormalities on a single case (satisfaction of search error), incorporates location information for each abnormality (right for wrong reason paradox), assigns greater weight to the most clinically significant abnormalities on the image and measures the confidence of the observers' decisions for each abnormality (uncertainty).[64] The summary measure of diagnostic accuracy compared to the reference standard diagnosis for free response studies is the figure of merit (FoM). A figure of merit was calculated for individual participants and for the average performance of the cohort of consultant radiologists and the cohort of reporting radiographers. Practitioner sensitivity was calculated at an abnormality level (number of abnormalities correctly identified) and specificity by the number of normal cases correctly identified.[279]

Both the Royal College of Radiologists and College of Radiographers are unwavering in their stance that any radiographer who undertakes image interpretation must perform at a level comparable to a consultant radiologist.[44] In recognition of this, and for this study, a non-inferiority approach was used, that is the study was designed to investigate if the reporting radiographers accuracy was *no worse* than the consultant radiologists (see Appendix 12.6). Utilising the non-inferiority approach, the null hypothesis that the FoM for the consultant radiologists exceeded the pre-defined clinical significance level (10% of average consultant radiologist) would be rejected if the p value was less than 0.05 using a one-tailed test, in line with CONSORT recommendations.[238]

4.3.1 Characteristics of the image bank

The image bank contained 106 cases, with an equal proportion of normal and abnormal cases (53 each). The number of lesions included in the 53 abnormal cases ranged from one to six, with an average of 2.28 lesions. A total of 121 lesions were included in the study.

A total of 21 reporting practitioners, ten consultant radiologists and eleven reporting radiographers, consented to participate in the study and completed the image bank. A total of 2,226 chest X-ray reports were expected to be generated. A small number of cases (total n=13) were not reported by some reporting practitioners, five (<1%) radiologists and eight (<1%) radiographers. There was no pattern to those missed, and the missed cases were attributed to participant oversight. A total of 2,213 reports were available for analysis; 1,055 (48%) by consultant radiologist and 1,158 (52%) by reporting radiographer.

4.3.2 Reporting practitioner sensitivity and specificity

A true positive (TP) was recorded if the reporting practitioner correctly identified and localised an abnormality according to the reference standard diagnosis and acceptance radius (Appendix 12.4.1.2). Any other abnormality identified was recorded as a false positive (FP). A true negative (TN) was recorded if the practitioner correctly identified a normal case as normal. The true positive and true negative results for the consultant radiologists and reporting radiographers is presented in Table 4.10.

Reporting Practitioner	True Positives (abnormalities)		True Negatives (cases)	
	Abnormalities (total)	Abnormalities (average)	Cases (total)	Cases (Average)
Consultant Radiologist	839 (total 1204)	83.9	428 (total 529)	42.8
Reporting Radiographer	1028 (total 1317)	93.5	491 (total 577)	44.6

Table 4.10 Consultant radiologist and reporting radiographer true positives and true negatives

Accounting for the small number of cases that were not interpreted by some practitioners, the consultant radiologists correctly identified 839 of 1,204 abnormalities and 428 of 529 normal cases. The reporting radiographers correctly identified 1,028 of 1,317 abnormalities and 491 of 577 normal cases.

Sensitivity was calculated at an abnormality/lesion level (proportion of abnormalities correctly identified). Specificity was calculated at a case level (correctly identifying a normal case as normal). The overall sensitivity and specificity of the consultant radiologists and reporting radiographers are presented in Figure 4.2.

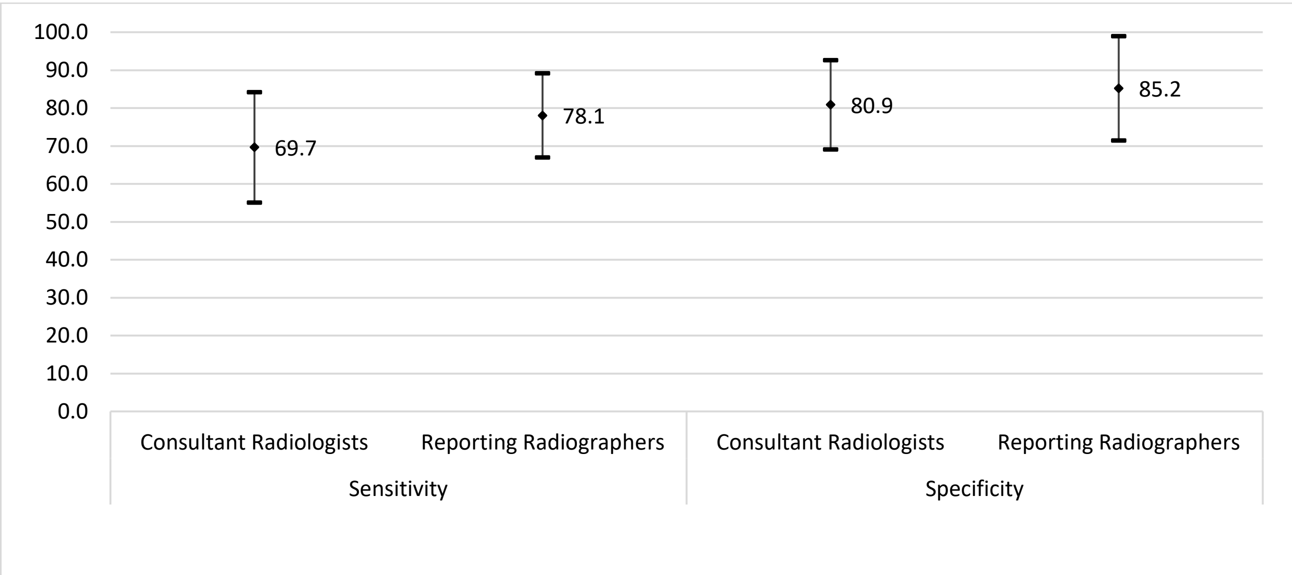


Figure 4.2 Sensitivity and specificity of consultant radiologists and reporting radiographers (with 95% CIs)

The consultant radiologists interpreted the image bank with 69.7% sensitivity (95%CI 55.1 - 84–2; range 54.6 – 81) and 80.9% specificity (95%CI 69.1 – 92.7; range 69.8 – 86.8). Reporting radiographer sensitivity (78.1% 95%CI 67 – 89.2; range 67.3 – 86) and specificity (85.2 95%CI 71.5 – 99; range 73.6 – 90.6) was broadly comparable.

4.3.3 Jack-knife alternate free response receiver operator characteristic curve (JAFROC) analysis

Alternate free response ROC studies require participants to locate abnormalities on the images and to assign a relative confidence score to each abnormality, accurately reflecting diagnostic decisions made in clinical practice (Appendix 12.4.1.2). Analysis of the data was performed using the jack-knife method and following best practice guidance.[241 246 249] Pseudovalues were calculated using JAFROC software (Chakraborty, version 4.2, August 2014)[257] and analysed using SPSS (IBM, version 21).

Unweighted JAFROC analysis assigns equal weight to each abnormality on an abnormal image. Weighted JAFROC analysis recognises that abnormalities on the images potentially have different clinical significance. For example, in a case which contains a malignant lung nodule and the patient being investigated has chronic obstructive pulmonary disease, detection of the malignant pathology will have a greater impact on patient management. In weighted analysis, the malignant lesions would be assigned a higher weighting.

All abnormal cases included in the study were reviewed by a professor of respiratory medicine and a consultant chest radiologist with a comprehensive case note summary and all imaging (previous and follow up) available. Weights were assigned to each lesion when multiple abnormalities were contained on a single case according to impact on patient management, based on a consensus decision between the professor of medicine and an expert chest radiologist (see Chapter 3.3.2 and Appendix 17). For example, Case 1 in the current study demonstrated cardiomegaly (relative weight assigned 0.98) and avascular necrosis of the spine (relative weight assigned 0.02) in a patient with dropping oxygen saturation and known sickle cell anaemia. Case 4 contained three calcified pleural plaques (relative weight assigned all 0.33) from previous asbestosis exposure in a patient with chest pain; and Case 39 demonstrated a left upper zone nodule (relative weight assigned 0.65) and a large left pleural effusion (relative weight assigned 0.35) in a patient who was later confirmed to have lung cancer.

The JAFROC performance, in terms of FoM values, of the consultant radiologists and reporting radiographers are presented in Table 4.11, Figures 4.3 and 4.4. The diagnostic accuracy of the cohort of reporting radiographers was non-inferior to that of the consultant radiologists for both unweighted JAFROC (t=11.826, p<0.0001) and weighted JAFROC (t=12.654, p<0.0001) analyses.

Reporting Practitioner	Number of Cases	Figure of Merit (95% CI)	
		Unweighted	Weighted
Consultant Radiologist	1055	0.788 (0.766 – 0.811)	0.786 (0.764 – 0.808)
Reporting Radiographer	1158	0.828 (0.808 – 0.847)	0.830 (0.811 – 0.849)

Table 4.11 Diagnostic Accuracy Figure of Merit Values of Consultant Radiologists and Reporting Radiographers

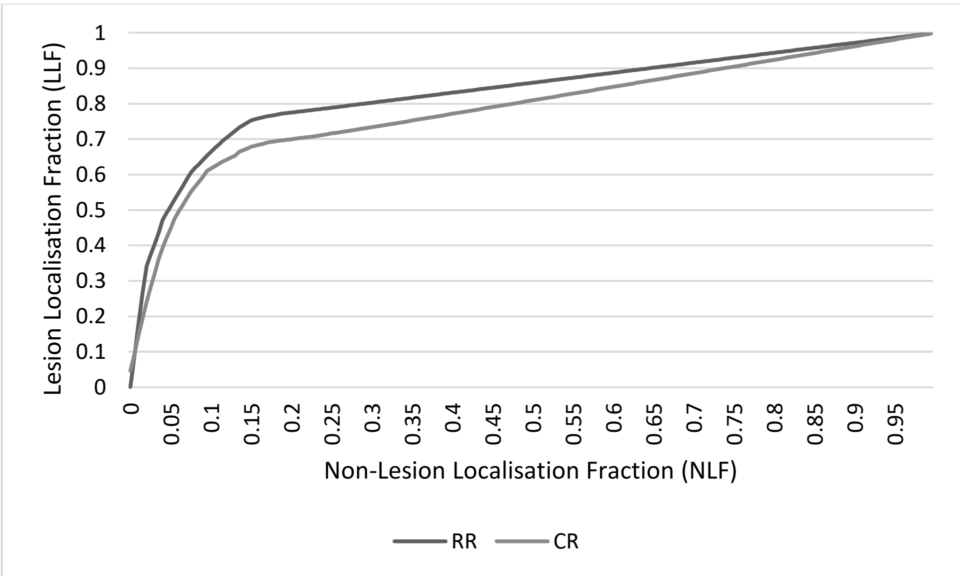


Figure 4.3 Unweighted JAFROC curves for consultant radiologists and reporting radiographers
CR = consultant radiologist; RR = reporting radiographer

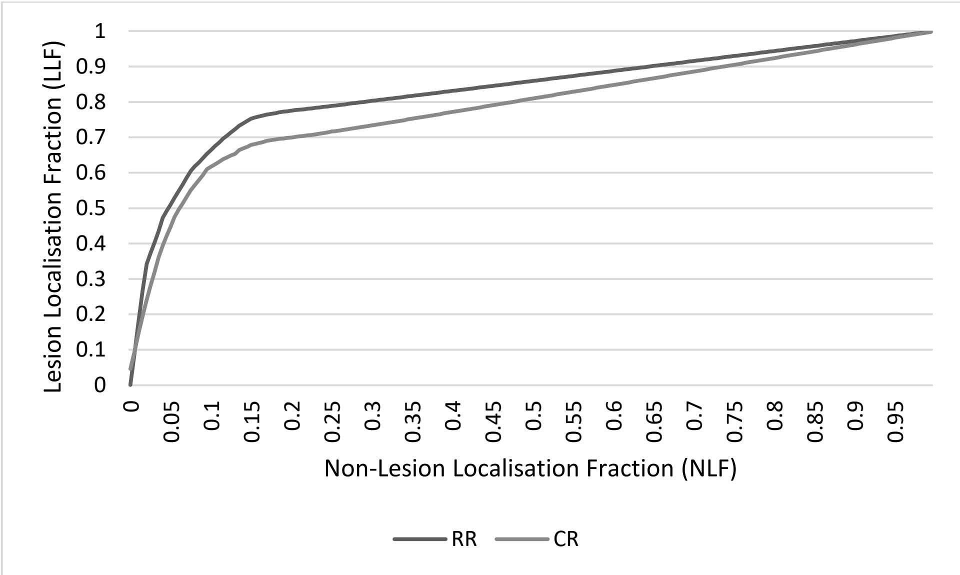


Figure 4.4 Weighted JAFROC curves for consultant radiologists and reporting radiographers
CR = consultant radiologist; RR = reporting radiographer

4.3.4 Influence of practitioner experience and volume of cases interpreted annually on performance

Practitioners' experience[82 228] and the number of cases reported annually[75] could influence diagnostic accuracy when interpreting chest X-rays. Participant demographics are presented in Table 4.12, with data from one practitioner missing.

	Consultant Radiologists			Reporting Radiographers		
	Volume			Volume		
Experience	< 5,000	5,001 – 9,999	≥ 10,000	< 5,000	5,001 – 9,999	≥ 10,000
0 – 5 years	2	^	^	1	4	1
6 – 9 years	4	^	^	1	2	1
≥ 10 years	2	2	^	^	^	^

Table 4.12 Participant experience and volume of chest X-rays reported annually
^ = no participant (Data from one reporting radiographer missing)

The diagnostic accuracy of the consultant radiologists and reporting radiographers, stratified by practitioner experience, are presented in Table 4.13.

	Weighted JAFROC Figure of Merit (95% CI)		
	Cons Rad	Rep Rad	Overall
Experience (years)			*
0 – 5	0.809 (0.750 – 0.881)	0.827 (0.777 – 0.877)	0.823 (0.781 – 0.864)
6 – 9	0.760 (0.710 – 0.811)	0.830 (0.777 – 0.883)	0.813 (0.771 – 0.855)
≥ 10	0.800 (0.750 – 0.850)	^	0.800 (0.750 – 0.850)

Table 4.13 Diagnostic accuracy of practitioners according to experience
Cons Rad = Consultant Radiologist, Rep Rad = Reporting Radiographer;
^ = no participant, *p=0.883

The experience of the reporting practitioners did not appear to influence diagnostic accuracy (F=0.125, p=0.883) when interpreting chest X-rays for either consultant radiologists or reporting radiographers.

The diagnostic accuracy of the consultant radiologists and reporting radiographers, stratified by the number of chest X-rays interpreted annually, are presented in Table 4.14. The number of chest X-rays interpreted annually by each practitioner did not appear to influence diagnostic accuracy (F=0.444, p=0.653) when interpreting chest X-rays.

	Weighted JAFROC Figure of Merit (95% CI)		
	Cons Rad	Rep Rad	Overall
Annual Volume			*
< 5,000	0.785 (0.748 – 0.823)	0.832 (0.760 – 0.903)	0.804 (0.768 – 0.840)
5,001 – 9,999	0.813 (0.742 – 0.884)	0.842 (0.798 – 0.885)	0.832 (0.794 – 0.870)
≥ 10,000	^	0.813 (0.742 – 0.884)	0.813 (0.742 – 0.884)

Table 4.14 Diagnostic accuracy of practitioners according to volume of chest X-rays interpreted
Cons Rad = Consultant Radiologist, Rep Rad = Reporting Radiographer;
^ = no participant, *p=0.653

The diagnostic accuracy of consultant radiologists and reporting radiographers, stratified by both experience and number of chest X-rays interpreted annually, is presented in Table 4.15.

	Consultant Radiologists			Reporting Radiographers		
	Volume			Volume		
Experience	< 5,000	5,001 – 9,999	≥ 10,000	< 5,000	5,001 – 9,999	≥ 10,000
0 – 5 years	0.809	^	^	0.839	0.839	0.803
6 – 9 years	0.760	^	^	0.824	0.844	0.822
≥ 10 years	0.787	0.813	^	^	^	^

Table 4.15 Diagnostic accuracy of practitioners according to experience and volume of chest X-rays interpreted
Cons Rad = Consultant Radiologist, Rep Rad = Reporting Radiographer; ^ = no participant

There was no apparent difference in diagnostic accuracy between the consultant radiologists and reporting radiographers for different current annual workload and experience. The small number of participants in each subset prevent further statistical analysis. This prevents firm conclusions from being drawn.

Radiographer reporting of chest X-rays is a relatively new practice, with the first programme of postgraduate education accredited in 2002. It may be that the current volume of cases reported each year by the radiographers balanced the experience of the consultant radiologists when interpreting chest X-rays. It is difficult to draw firm conclusions due to the small number of practitioners in each subcategory.

4.3.5 Diagnostic accuracy of individual practitioners – JAFROC

Figure of merit values for each individual practitioner for both unweighted and weighted JAFROC were calculated and are presented in Table 4.16. The individual reporting practitioners were ranked by performance by the primary outcome measure (weighted JAFROC).

Rank	Reporting Practitioner	Weighted	Unweighted
1	Rep Rad	0.905	0.902
2	Rep Rad	0.860	0.838
3	Rep Rad	0.848	0.867
4	Cons Rad	0.842	0.846
5	Rep Rad	0.839	0.819
6	Rep Rad	0.829	0.840*
7	Cons Rad	0.827	0.832*
8	Cons Rad	0.827	0.817
9	Rep Rad	0.824	0.817
10	Rep Rad	0.822	0.817
11	Rep Rad	0.819	0.807
12	Cons Rad	0.805	0.819*
13	Rep Rad	0.803	0.805*
14	Rep Rad	0.799	0.801*
15	Cons Rad	0.792	0.789
16	Cons Rad	0.784	0.791*
17	Rep Rad	0.783	0.789*
18	Cons Rad	0.783	0.793*
19	Cons Rad	0.769	0.778*
20	Cons Rad	0.716	0.704
21	Cons Rad	0.715	0.713

Table 4.16 Diagnostic accuracy of individual reporting practitioners
Rep Rad = Reporting Radiographer, Cons Rad = Consultant Radiologist,
* = unweighted higher than weighted (9 practitioners, 5 Cons Rad, 4 Rep Rad)

Individual diagnostic accuracy ranged from 0.715 – 0.905 for weighted JAFROC and 0.713 – 0.902 for unweighted JAFROC. Of the top five practitioners with the highest diagnostic accuracy for both weighted and unweighted JAFROC, four were reporting radiographers. Four of the practitioners with the lowest diagnostic accuracy were consultant radiologists.

The diagnostic accuracy, in terms of FoM, was higher for most practitioners for weighted compared to unweighted JAFROC. This suggests that in general, both reporting radiographers and consultant radiologists had a tendency to identify the most clinically relevant abnormality on abnormal images and when misses occurred, these tended to be the less relevant abnormalities. For example, all participants correctly identified a case of confirmed lung cancer but several missed the small pleural effusion (Case 101; all identified the primary malignant lesion, but only 55% of radiographers (n=6) and 50% radiologists (n=5) identified the small pleural effusion).

4.3.6 Inferred receiver operator characteristic curve (ROC) analyses to facilitate comparison with data from other sources

Traditional receiver operator characteristic (ROC) methodology utilises a binary decision whereby the observer rates the case as either normal or abnormal without localisation information. For many years, this was the standard methodology for assessing diagnostic accuracy, but this methodology has several important limitations as was discussed in Appendix 12.4.1.1.

Although the free response methodology paradigm, of which JAFROC is the most established and modern derivation, has been used in diagnostic accuracy studies since 2005, some authors continue to use ROC. So to facilitate comparison with the existing evidence base examining chest X-ray interpretation, inferred ROC was calculated using JAFROC 4.2.[257] For inferred ROC, the highest-ranking lesion on abnormal cases is taken as the inferred ROC and standard area under the curve (AUC) values are calculated. For this study, these are presented in Table 4.17 and Figure 4.5.

Reporting Practitioner	Area Under the Curve (AUC) (95% confidence intervals)
Consultant Radiologists	0.903 (0.882 – 0.924)
Reporting Radiographers	0.909 (0.887 – 0.931)

Table 4.17 Diagnostic Accuracy of consultant radiologists and reporting radiographers – Inferred ROC

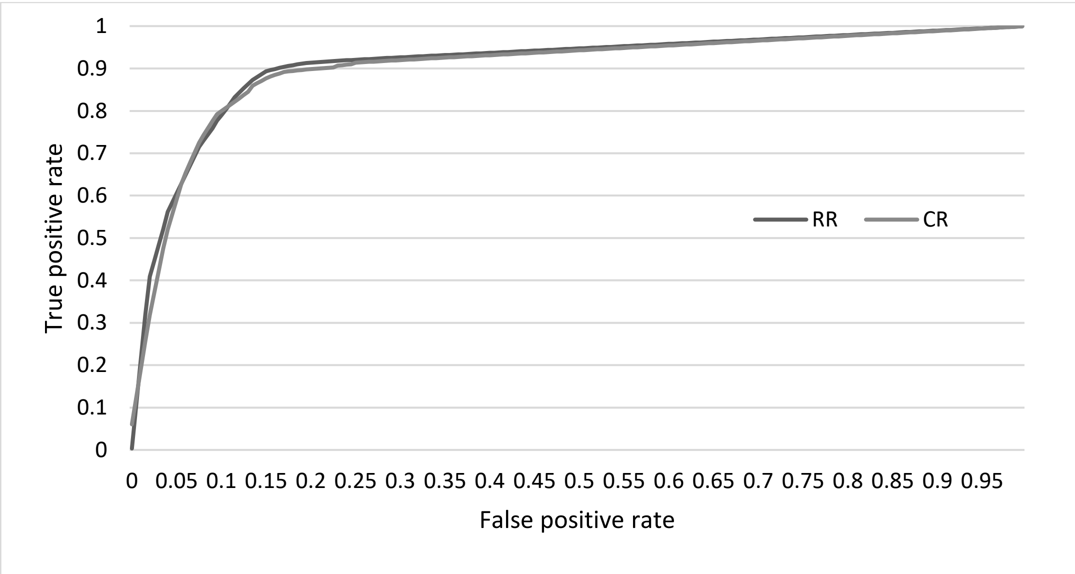


Figure 4.5 Inferred receiver operator characteristic curves for consultant radiologists and reporting radiographers
CR = consultant radiologist; RR = reporting radiographer

The diagnostic accuracy of the cohort of reporting radiographers (AUC 0.909; 95%CI 0.887 – 0.924) was non-inferior to that of the consultant radiologists (0.903; 95%CI 0.887 – 0.931) (t=9.610, p<0.0001).

4.3.7 Diagnostic accuracy of individual practitioners – inferred ROC

Individual diagnostic accuracy for inferred ROC was calculated and area under the curve results are ranked and presented in Table 4.18.

Rank	Reporting Practitioner	Inferred ROC (AUC)
1	Rep Rad	0.972
2	Rep Rad	0.960
3	Cons Rad	0.936
4	Cons Rad	0.935
5	Cons Rad	0.932
6	Cons Rad	0.919
7	Rep Rad	0.914
8	Rep Rad	0.914
9	Rep Rad	0.908
10	Cons Rad	0.908
11	Rep Rad	0.905
12	Cons Rad	0.906
13	Rep Rad	0.901
14	Rep Rad	0.898
15	Cons Rad	0.897
16	Rep Rad	0.894
17	Rep Rad	0.881
18	Cons Rad	0.873
19	Cons Rad	0.869
20	Cons Rad	0.857
21	Rep Rad	0.853

Table 4.18 Inferred ROC diagnostic accuracy for individual practitioners
Rep Rad = Reporting Radiographer, Cons Rad = Consultant Radiologist

Individual diagnostic accuracy ranged from 0.853 – 0.972 for inferred ROC. Of the top five practitioners with the highest diagnostic accuracy for inferred ROC, three were consultant radiologists. Three of the practitioners from the lowest ranking five were consultant radiologists, although the lowest ranked reporting practitioner was a reporting radiographer.

4.3.8 Comparison between JAFROC and inferred ROC analysis

Comparison was made between the diagnostic accuracy of practitioners for the two main analysis methods used in this study; weighted JAFROC and inferred ROC (Appendix 12.4.1.2 and Chapter 4.3.5). The relative performance of the reporting radiographers and consultant radiologists are presented in Table 4.19.

Rank	Weighted JAFROC (Practitioner ID)	Practitioner Group	Inferred ROC (Practitioner ID)	Practitioner Group
1	2	Rep Rad	2	Rep Rad
2	24	Rep Rad	6	Rep Rad
3	6	Rep Rad	23	Cons Rad
4	21	Cons Rad	21	Cons Rad
5	3	Rep Rad	10	Cons Rad
6	16	Rep Rad	9	Cons Rad
7	10	Cons Rad	24	Rep Rad
8	7	Cons Rad	16	Rep Rad
9	4	Rep Rad	12	Rep Rad
10	1	Rep Rad	19	Cons Rad
11	17	Rep Rad	1	Rep Rad
12	19	Cons Rad	7	Cons Rad
13	8	Rep Rad	17	Rep Rad
14	22	Rep Rad	3	Rep Rad
15	18	Cons Rad	13	Cons Rad
16	23	Cons Rad	4	Rep Rad
17	12	Rep Rad	22	Rep Rad
18	13	Cons Rad	20	Cons Rad
19	9	Cons Rad	5	Cons Rad
20	5	Cons Rad	18	Cons Rad
21	20	Cons Rad	8	Rep Rad

Table 4.19 Individual practitioner diagnostic accuracy for weighted JAFROC and inferred ROC
Rep Rad = Reporting Radiographer, Cons Rad = Consultant Radiologist

The majority of practitioners had similar performance regardless of the methodology used to determine diagnostic accuracy. Participant 2 (reporting radiographer) had the highest diagnostic accuracy in both analyses. There are however some variations; for example

Participant 8 (reporting radiographer) was ranked 13th when JAFROC was used (FoM = 0.803) but ranked bottom using the inferred ROC (AUC = 0.853).

4.3.9 Abnormality detection by pathology

To demonstrate whether performance between consultant radiologists and reporting radiographers were comparable for different pathologies, analysis of abnormality detection was performed and stratified by pathology.

4.3.9.1 True positive abnormalities

Abnormalities correctly identified by each reporting practitioner were determined and grouped by pathological category and practitioner group. A summary of the pathologies correctly identified by both consultant radiologists and reporting radiographers is shown in Table 4.20.

True Positive				Number of Occasions	
Pathology	Feature			Observer	
				RR	CR
Cardiac	Cardiomegaly			31	26
	Pericardial Effusion			21	30
	Pulmonary Oedema			69	52
	Interstitial Oedema	Left		58	39
		Right		62	40
Pleural	Effusion	Unilateral		105	95
		Bilateral		38	25
	Pneumothorax			10	8
	Plaque			32	27
Mediastinal	Hilum	Vascular	Unilateral	4	3
			Bilateral	2	0
		Lymphadenopathy	Unilateral	3	2
			Bilateral	1	0
	Paratracheal			13	17
Infection	Consolidation	Left Lung		80	68
		Right Lung		117	97
	Atelectasis			25	20
Malignant	Whiteout	Left Lung		11	9
	Nodule	Left		22	20
		Right		27	24
	Metastases	Bilat Lower Zones		9	7
	Myeloma	All 3 Lesions		0	2
		1 or 2 Lesions		7	5
Airways Disease	Chronic Obstructive Pulmonary Disease			40	30
	Fibrosis	Left		45	35
		Right		30	25
Skeletal	Fracture or Dislocation			12	7
	Avascular Necrosis			1	1
Lines & Tubes	nasogastric tube mis-sited			22	19

Table 4.20 Nature of the true positive lesions identified by each practitioner group. RR = Reporting Radiographer, CR = Consultant Radiologist.

In addition to overall measures of performance, such as figures of merit (JAFROC) and area under the curve (ROC) it is important to ensure that pathologies with significant impact on patient outcome (for example malignancy, pneumothorax) are recognised. The consultant radiologists and reporting radiographers were broadly comparable for abnormalities correctly identified, with no overall trend seen.

Sub analysis of the malignant cases included in the image bank revealed similar performance between consultant radiologists and reporting radiographers. Three radiologists (of ten) identified all seven malignant cases compared to five reporting radiographers (of eleven). In addition, both groups of practitioners correctly identified the majority of abnormalities in each case. The lowest detection rate was for Case 107, a subtle right mid zone nodule with no other associated findings. This case was selected from the discrepancy meeting files and was correctly identified by eight (73%) of the reporting radiographers and six (60%) of the consultant radiologists. The single case of a pneumothorax was correctly identified by ten (of eleven, 91%) reporting radiographers and eight (of nine, 89%) consultant radiologists. Two cases included a mis-sited nasogastric tube, a significant abnormality with potentially serious consequences if not identified. All reporting radiographers identified both of these cases (n=22 observations), while one was not identified by a consultant radiologist (n=19 occasions).

4.3.9.2 False positive abnormalities

Abnormalities incorrectly identified by the reporting practitioners were also grouped by pathology type for analysis. A summary is presented in Table 4.21.

False Positives				Number of Occasions	
Pathology	**Feature**			**Observer**	
				RR	*CR*
Cardiac	Cardiomegaly			71	57
	Upper Lobe Blood Diversion			24	20
Pleural	Effusion	Unilateral		28	40
		Bilateral		14	10
	Plaque			6	4
	Pneumothorax			0	3
Mediastinum	Hilum	Vascular	Unilateral	8	8
			Bilateral	10	9
		Lymphadenopathy	Unilateral	10	3
			Bilateral	2	11
	Paratracheal	Left		2	0
		Right		2	5
Infection	Consolidation	Unilateral		52	61
		Bilateral		12	5
	Atelectasis	Left		4	16
		Right		15	12
Malignant	Nodule	Left		8	14
		Right		9	15
Tubes & lines		nasogastric tube mis-sited		4	2
		internal jugular central line mis-sited		2	0
		implantable cardiac defibrillator lead mis-sited		1	0
		endotracheal tube mis-sited		0	1
Airways Disease	Chronic Obstructive Pulmonary Disease			40	32
	Interstitial	Left		7	14
		Right		8	13
	Bronchiectasis			5	5
Skeletal	Fracture			9	11
	Avascular Necrosis			3	6
Abdominal	Dilated Bowel			3	2
	Perforation			1	1

Table 4.21 Nature of the false positive errors made by each practitioner group. RR = Reporting Radiographer, CR = Consultant Radiologist.

No trends were identified in the false positive errors of the consultant radiologists and the reporting radiographers. Cardiomegaly was more incorrectly diagnosed by the reporting radiographers, 71 occurrences compared to 57. Consultant radiologists more often incorrectly diagnosed a malignant lung nodule, 29 false positive errors compared to 17 by the reporting radiographers.

4.3.9.3 False negative abnormalities

The summary of the false negative errors made by the consultant radiologists and reporting radiographers by pathology type are presented in Table 4.22.

False Negative				Number of Occasions	
Pathology	**Feature**			**Observer**	
				RR	*CR*
Cardiac	Cardiomegaly			0	4
Pleural	Effusion	Unilateral		36	32
		Bilateral		13	17
	Pneumothorax			1	1
	Plaque			1	1
Mediastinal	Hilum	Vascular	Unilateral	1	4
			Bilateral	5	4
		Lymphadenopathy	Unilateral	1	0
			Bilateral	7	7
	Paratracheal	Right		9	2
Infection	Consolidation	Left		11	17
		Right		3	13
	Atelectasis			3	8
	Cavity			12	4
Malignant	Nodule	Left		0	1
		Right		6	5
	Metastases	Bilat Lower Zones		1	2
Airways Disease	Chronic Obstructive Pulmonary Disease			4	9
	Interstitial	Left		39	52
		Right		15	18
Skeletal	Dislocation	Right Shoulder		9	9
	Avascular Necrosis	Spine		10	8
	Myeloma	Ribs		25	16

Table 4.22 Nature of the false negative errors by each practitioner group.
RR = Reporting Radiographer, CR = Consultant Radiologist.

The false negative errors made by each practitioner group were broadly similar. Pleural effusions were missed on 49 occasions by both the consultant radiologists and reporting radiographers. Consultant radiologists did not identify malignant lung nodules on eight occasions compared to seven reporting radiographer misses. Lymphadenopathy was missed in seven instances by consultant radiologists and eight by reporting radiographers.

Cardiomegaly was more likely to be identified by the reporting radiographers than the consultant radiologists, both correctly (true positive 31 vs. 26 occasions) and incorrectly (false positive 71 vs. 57 occasions). Consultant radiologists correctly identified fewer pleural effusions (95 unilateral and 25 bilateral) than the reporting radiographers (105 unilateral and 38 bilateral), but incorrectly diagnosed effusions on 50 occasions (40 unilateral, 10 bilateral) compared to 42 radiographer incorrect diagnoses (28 unilateral, 14 bilateral). There were no cases of apical pleural thickening outside of other related pathologies (post-surgical change, asbestos related pleural disease). This was incorrectly identified on 13 occasions by the radiologists and once by the radiographers. Interstitial oedema was more frequently correctly identified by the cohort of reporting radiographers compared to the consultant radiologists, 120 and 79 instances respectively.

4.3.10 Summary of performance in reporting accuracy by consultant radiologists and reporting radiographers

Generally, there was good agreement and similar level of reporting accuracy for both consultant radiologists and reporting radiographers as detailed in the sections above. For all measures of diagnostic accuracy; unweighted JAFROC (CR=0.788; RR=0.828) weighted JAFROC (CR=0.786; RR=0.830) and inferred ROC (CR=0.903; RR=0.909) the performance of the reporting radiographers was non-inferior to that of the consultant radiologists. There were no apparent differences when the true positive, false positive and false negative decisions were examined by pathology type.

4.4 Part 2 – Influence of chest X-ray reports on clinicians' diagnostic decision-making

A number of factors can influence clinicians' diagnostic decision-making. As detailed in Chapter 2.4, these include the patient's history, previous investigations and, pertinent to this work, radiology investigations and reports. For any medical investigation to alter patient management the results of the test must first influence the diagnostic decision-making of the treating clinician, by either suggesting a new diagnosis and/or altering the confidence related to an existing diagnosis.

The consultant radiologist and reporting radiographer chest X-ray reports produced in the diagnostic accuracy study (Chapter 3.2.8 and Chapter 4.3) were given to a cohort of clinicians. The clinicians were asked to indicate a most likely and most serious diagnosis for each case and to assign a confidence rating for each diagnosis, both prior to (pre-CXR diagnosis) and then in conjunction with a chest X-ray report (post-CXR diagnosis). The pre-CXR and post-CXR diagnosis was compared to the clinico-radiological (final) diagnosis (Chapter 3.3.2) for accuracy. The aim of this part of the study was to demonstrate whether the chest X-ray reports of consultant radiologists and reporting radiographers have different influences on the diagnostic decision-making of clinicians.

> Research Question: Is there a clinically significant difference between the influence that consultant radiologist and reporting radiographer chest X-ray reports have on clinicians' diagnostic decision-making?
>
> Alternative Hypothesis [non-inferiority approach]: There is no clinically significant difference in the influence between chest X-ray reports produced by consultant radiologists or by reporting radiographers have on clinicians' diagnostic decision-making

Null Hypothesis [non-inferiority approach]: Consultant radiologists' chest X-ray reports will have a clinically significant difference on clinicians' diagnostic decision-making when compared to reporting radiographers' chest X-ray reports.

The first part of this section summarises the clinico-radiological (final) diagnosis for the cases included in the study. The influence of chest X-ray reports on the clinicians' most likely and most serious diagnosis are examined, with accuracy of the diagnosis and alterations in diagnostic confidence the primary outcome measures. Comparisons are made between consultant radiologist and reporting radiographer chest X-ray reports, and differences between different diagnoses and for clinicians of different experience are also examined. Changes in diagnostic confidence with the chest X-ray reports are presented, both uncorrected and corrected according to the accuracy of the diagnosis utilising the Tsushima methodology[277] (Chapter 3.3.6).

4.4.1 Characteristics of the image bank used by the clinicians

A professor of respiratory medicine performed a comprehensive case note review for all 106 cases included in the study. This was undertaken between 12 and 24 months following the acquisition of the chest X-ray in clinical practice. For the purposes of this study, the review and conclusion of this senior clinician was taken as the final clinico-radiological diagnosis for each case. A concise summary of the salient clinical findings and features (patient history, laboratory results, physical findings; see Chapter 3.2.5) available at the time of the original clinical investigation was also produced. All case notes were available for review (n=106, 100%) by the respiratory professor, and a summary of the final diagnoses for all cases are summarised in Table 4.23.

Diagnosis	Frequency	Percentage
No Significant Disease	36	34
Infection	18	17
Cardiac/Pulmonary Oedema	15	14
Pleural Effusion	1	1
Chronic Obstructive Pulmonary Disease	4	4
Pneumothorax	2	2
Perforation	0	0
Malignancy	9	8
Tuberculosis	6	6
Pulmonary Fibrosis	2	2
Other	13	12
TOTAL	106	100

Table 4.23 Final clinico-radiological diagnosis of the cases included in the study

The most frequent final diagnosis was no significant disease (36 cases, 34%). Infection (18 cases, 17%), cardiac/pulmonary oedema (15 cases, 14%) and malignancy (9 cases, 8 %) the three most frequent abnormal final clinico-radiological diagnoses in the case series. A range of other respiratory pathologies each provided a small contribution. Perforation was included in the list of possible differential diagnoses provided to the clinicians although there were no occurrences in the case series.

4.4.2 Characteristics of participant clinicians [experience, specialities]

Of the 27 clinicians (Table 4.24) who were recruited to the study and consented to participate, one participant completed the pre chest X-ray report diagnoses with the remaining participants (8, 30%) withdrawing prior to data collection. Fourteen clinicians (52%) completed all case summaries (all pre and post diagnoses) and four clinicians (15%) partially completed the post chest X-ray report diagnoses. A total of 18 clinicians participated therefore in this part of the study. Clinicians who provided partial responses (n=4) were included.

The process of making diagnostic decisions has been shown to change as clinicians gain experience (Chapter 2.4[280]). To replicate clinical practice and to examine for potential differences in how a chest X-ray report may influence diagnostic decision-making, clinicians with a range of experience were recruited to the study (Chapter 3.3.5).

Participant Number	Grade	Speciality	Completed
1	Junior Medical	Respiratory	Full
2	Junior Medical	Respiratory	Withdrew
3	Registrar	Respiratory	Full
4	Registrar	Respiratory	Full
5	Junior Medical	Respiratory	Withdrew
6	Consultant	Respiratory	Full
7	Junior Medical	General Medicine	Full
8	Registrar	Respiratory	Partial – all pre CXR only
9	Consultant	General Medicine	Full
10	Consultant	Respiratory	Partial – all pre & some post CXR
11	Consultant	Emergency Medicine	Full
12	Registrar	Emergency Medicine	Withdrew
13	Registrar	Respiratory	Full
14	Consultant	General Medicine	Full
15	Junior Medical	General Medicine	Full
16	Junior Medical	General Medicine	Full
17	Registrar	General Medicine	Withdrew
18	Consultant	Care of the Elderly	Full
19	Consultant	Care of the Elderly	Partial – all pre & some post CXR
20	Consultant	General Medicine	Withdrew
21	Consultant	Respiratory	Withdrew
22	Junior Medical	General Medicine	Partial – all pre & some post CXR
23	Registrar	Respiratory	Full
24	Junior Medical	General Medicine	Withdrew
25	Junior Medical	General Medicine	Withdrew
26	Junior Medical	General Medicine	Full
27	Registrar	Respiratory	Partial – all pre & some post CXR

Table 4.24 Demographics of clinician participants and completion rates

4.4.3 Influence of chest X-ray reports on clinicians diagnostic decision-making

As outlined in the literature (Chapter 2.5 and Chapter 3.3), when the chest X-ray report is available and supports the clinicians in a correct diagnosis, both for a new diagnosis and a retained initial diagnosis, the report appeared to result in a beneficial impact on the clinicians' diagnostic decision-making. Where an incorrect diagnosis is reached by the clinician both prior to and then with the chest X-ray report, the influence of the report is neutral. Where a correct initial diagnosis is changed to an incorrect diagnosis with the chest X-ray report, the report had a detrimental influence on the clinicians' diagnostic decision-making. A summary of the benefit or harm of chest X-ray reports is demonstrated in Figure 4.6. This data is detailed in the following sections.

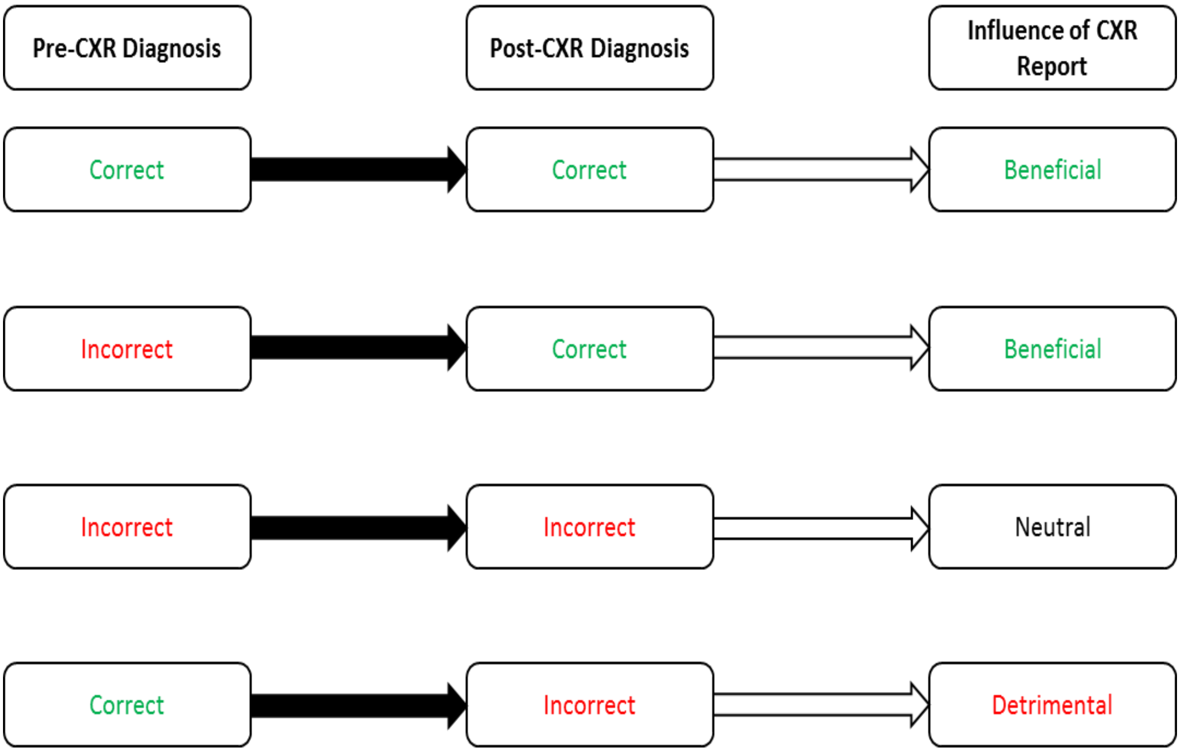


Figure 4.6 Influence of a chest X-ray report on clinicians' diagnostic decision-making (adapted from Tsushima *et al.* 2003[277])

4.4.4 Clinicians' confidence in diagnoses

For all cases included in the study, the clinicians were required to assign a confidence level to each most likely and each most serious diagnosis using a continuous scale (0 = not considered or very unlikely – 100 = definite) in line with established conventions and previous research.[163 277] Firstly, they were required to do this based on the case summary alone (pre-CXR confidence) and, secondly, with the case summary and the chest X-ray report (post-CXR confidence) after a washout period (Chapter 2.5 and Chapter 3.3.5). Changes in the diagnostic confidence of clinicians were determined by comparing the pre-CXR and post-CXR confidences (Chapter 3.3.5). Analysis was performed using both uncorrected and corrected (post-CXR diagnosis compared to final clinico-radiological diagnosis) diagnostic confidences. Uncorrected diagnostic confidence was determined by comparing the pre-CXR and post-CXR diagnostic confidence of the clinicians irrespective of the accuracy of the post-CXR diagnosis.

The accuracy of a clinician's diagnosis is fundamental to appropriate prognostic and therapeutic decisions; it is has been suggested that there is nothing more dangerous than a clinician who is confidently wrong.[275] The corrected diagnostic confidence scores for both the most likely and most serious diagnoses of clinicians for consultant radiologist and reporting radiographer chest X-ray reports were compared. Correction of the diagnostic confidence was performed using the Tsushima method where -200 (correct pre-CXR diagnosis with 100% confidence changed to an incorrect diagnosis with 100% confidence with the chest X-ray report) to +200 (incorrect pre-CXR diagnosis with 100% confidence changed to a correct post-CXR diagnosis with 100% confidence).

4.4.5 Clinician most likely diagnosis

Clinicians' experience (Table 4.25) did not appear to have an influence on their diagnostic decisions.

Pre-CXR Most Likely Diagnosis	Consultants (n=7)		Registrars (n=5)		Junior Medical Staff (n=6)		Total		
	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	**All**
No Significant Disease	315	409	227	272	152	357	694	1038	**1732**
Infection	196	165	161	118	251	175	608	458	**1066**
Cardiac/Pulmonary Oedema	67	51	13	67	47	49	127	167	**294**
Pleural Effusion	0	0	0	0	6	0	6	0	**6**
COPD	31	2	54	36	58	45	143	83	**226**
Pneumothorax	28	5	11	0	20	6	59	11	**70**
Perforation	20	0	0	0	18	0	38	0	**38**
Malignancy	21	34	23	48	9	31	53	113	**166**
Tuberculosis	80	67	42	11	28	51	150	129	**279**
Pulmonary Fibrosis	0	27	0	12	3	12	3	51	**54**
Other	173	7	140	15	189	40	502	62	**564**
TOTAL	931	767	671	579	781	766	2383	2112	**4495**

Table 4.25 Accuracy of clinician pre-CXR most likely diagnoses based on clinical case summary
COPD = chronic obstructive pulmonary disease

There was little variation in the proportion of correct decisions made (consultants 767 of 1,698 45%; registrars 579 of 1,250, 46%; junior medical staff 766 of 1,547, 50%).

4.4.5.1 <u>Influence of chest X-ray reports on clinicians' most likely diagnoses compared with diagnosis reached without supporting X-ray report</u>

After a minimum washout period of seven days, a statistically powered number of clinicians (n=18 total; n=4 partially completed, n=14 completed all summaries) were presented with the same clinical case summaries for each case and either a consultant radiologist or reporting radiographer chest X-ray report associated with the case. The post-CXR most likely diagnosis was compared to the initial (pre-CXR) diagnosis provided by the clinician and to the final clinico-radiological diagnosis for changes in diagnosis and for accuracy of the post-CXR diagnosis (Chapter 3.3.5). Clinicians were blinded to the reporting practitioner who provided the report.

Again, clinicians' experience did not appear to affect their clinical decisions (Table 4.26 and 4.27).

Clinician Experience	Most Likely Diagnosis	
	Retained	New
Consultant	808	856 (51%)
Registrar	606	619 (51%)
Junior Medical Staff	752	750 (50%)
Total	2166	2225 (51%)

Table 4.26 Chest X-ray reports that produced a change in clinician most likely post-CXR diagnosis for clinicians of different experience

Clinician Experience	Most Likely Diagnosis Consultant Radiologist		Most Likely Diagnosis Reporting Radiographer	
	Retained	New	Retained	New
Consultant	426	475 (53%)	382	381 (50%)
Registrar	320	279 (47%)	286	340 (54%)
Junior Medical Staff	344	334 (49%)	408	416 (50%)
Total	1090	1088 (50%)	1076	1137 (51%)

Table 4.27 Cases which produced a new most likely diagnosis post-CXR

4.4.5.2 Accuracy of most likely and/or most serious diagnosis

The most likely and/or most serious diagnoses reached by the clinicians of different experience with the case summary and chest X-ray report (post-CXR) were compared to the final clinico-radiological diagnosis for each case. The results are presented in Table 4.28.

Clinician Experience	Most Likely and/or Most Serious Diagnosis	
	Incorrect	Correct
Consultant	647	1017 (61%)
Registrar	485	740 (60%)
Junior Medical Staff	554	948 (63)
Total	1686	2705 (62%)

Table 4.28 Number of cases with a correct post-CXR diagnosis (most likely or most serious) for clinicians of different experience

The clinicians' experience did not appear to be important for a correct most likely or most serious diagnosis (Table 4.29). An accurate diagnosis was reached by clinicians in 2,705 (62%) of cases, with a similar proportion across the different clinician grades as noted in Table 4.28 (above).

Clinician Experience	Correct Most Likely and/or Most Serious Consultant Radiologist		Correct Most Likely and/or Most Serious Reporting Radiographer	
	Incorrect	Correct	Incorrect	Correct
Consultant	337	564 (63%)	310	453 (59%)*
Registrar	217	382 (64%)	268	358 (57%)**
Junior Medical Staff	256	422 (62%)	298	526 (64%)***
Total	810	1368 (63%)	876	1337 (60%)****

Table 4.29 Number of cases with a correct post-CXR diagnosis (most likely or most serious) for clinicians of different experience with consultant radiologist and reporting radiographer chest X-ray reports

Chi-square; *p=0.179; **p=0.018; *** p=0.524; **** p=0.103

There was a trend for correct diagnoses by clinicians (Table 4.29) to be associated with consultant radiologist reports than with reporting radiographers; 63% (1,368 of 2178) and 60% (1,337 of 2,213) respectively, but not statistically different (Chi-square χ=2.66, p=0.103).

4.4.5.3 Influence on clinicians' most likely diagnosis of chest X-ray reports from consultant radiologists and reporting radiographers

Cases in which consultant radiologist and reporting radiographer chest X-ray reports had a beneficial, neutral and detrimental influence (Figure 4.6) on clinicians' most likely diagnoses are presented in Table 4.30 and Figure 4.7.

Clinician Experience	Most Likely post-CXR diagnosis Consultant Radiologist				Most Likely post-CXR diagnosis Reporting Radiographer			
	Beneficial		Detrimental	Neutral	Beneficial		Detrimental	Neutral
	Retained Correct	New Correct	New Incorrect	Retained Incorrect	Retained Correct	New Correct	New Incorrect	Retained Incorrect
Consultant	140 (16%)	207 (23%)	269 (30%)	285 (32%)	105 (14%)	159 (21%)	236 (31%)	263 (34%)
Registrar	76 (13%)	95 (16%)	212 (35%)	216 (36%)	93 (15%)	108 (17%)	188 (30%)	237 (39%)
Junior Medical Staff	114 (17%)	113 (17%)	225 (33%)	226 (33%)	133 (16%)	148 (18%)	275 (33%)	268 (33%)
Total	330 (15%)	415 (19%)	706 (32%)	727 (34%)	331 (15%)	415 (19%)	699 (32%)	768 (34%)

Table 4.30 Influence of consultant radiologist and reporting radiographer chest X-ray reports on the most likely diagnoses of clinicians

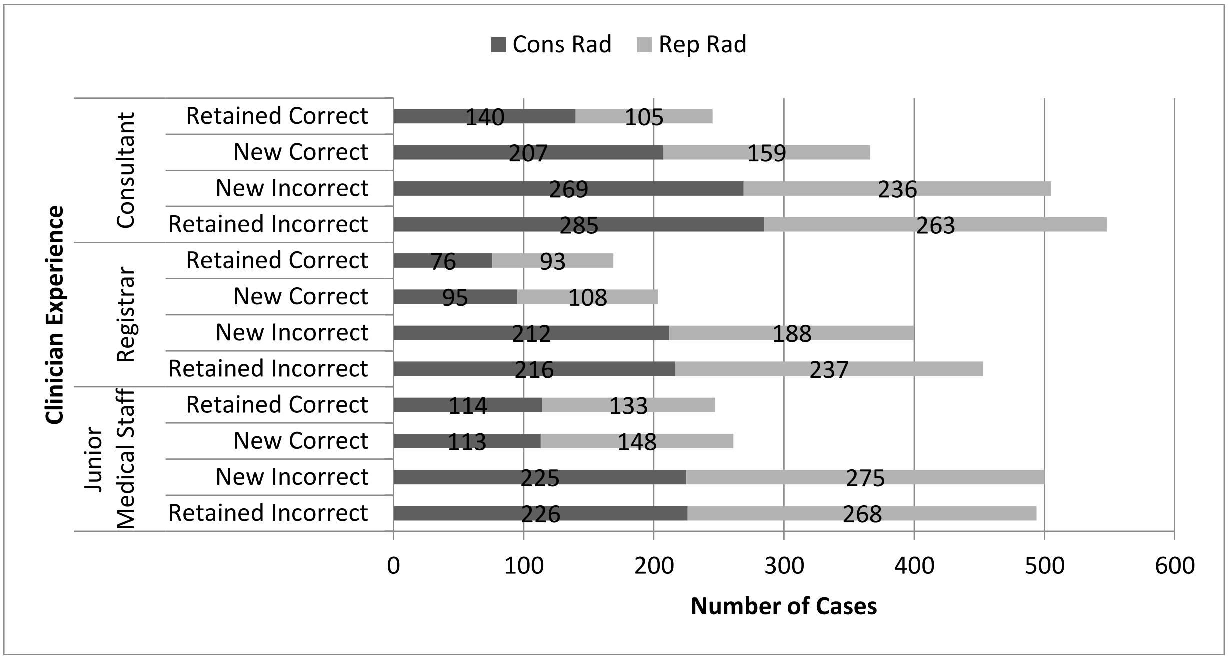


Figure 4.7 Influence of consultant radiologist and reporting radiographer chest X-ray reports on the most likely diagnoses of clinicians
Cons Rad = consultant radiologist; Rep Rad = reporting radiographer

In general, the clinicians' experience does not appear to have an influence on their diagnostic decisions. The origin of the report, whether consultant radiologist (415 of 2,178, 19%) or reporting radiographer (415 of 2,213, 19%) does not appear to influence clinicians' decisions either. There is a non-significant trend (Chi-square χ=0.882, p=0.83) indicating clinicians may prefer radiologists reports. No apparent difference was found between the influence on clinician most likely diagnosis for consultant radiologist and reporting radiographer chest X-ray reports when the experience of the clinician was considered (consultant χ=2.91, p=0.406; registrar χ=4.363, p=0.225; junior medical staff χ=0.539, p=0.91)

4.4.5.4 Consultant radiologist and reporting radiographer chest X-ray reports influence on clinicians' most likely diagnosis by final clinico-radiological diagnosis

The influence that consultant radiologist and reporting radiographer chest X-ray reports had on the most likely diagnosis of clinicians' is shown in Table 4.31.

Final Diagnosis (case note review)	Most Likely post-CXR diagnosis Consultant Radiologist				Most Likely post-CXR diagnosis Reporting Radiographer				Chi-square (p value)
	Beneficial		Detrimental	Neutral	Beneficial		Detrimental	Neutral	
	Retained Correct	New Correct	New Incorrect	Retained Incorrect	Retained Correct	New Correct	New Incorrect	Retained Incorrect	
No Significant Disease	121 (14%)	181 (22%)	387 (46%)	147 (18%)	126 (15%)	152 (19%)	385 (47%)	150 (18%)	2.342 (p=0.505)
Infection	64 (15%)	93 (22%)	162 (38%)	105 (25%)	56 (13%)	106 (25%)	166 (40%)	90 (22%)	2.543 (p=0.468)
Cardiac/ Pulmonary Oedema	37 (16%)	23 (10%)	42 (18%)	128 (56%)	50 (19%)	26 (10%)	36 (14%)	150 (57%)	2.257 (p=0.521)
Pleural Effusion	^	7 (27%)	^	19 (73%)	^	5 (21%)	^	19 (79%)	0.254 (p=0.614)
COPD	12 (24%)	4 (8%)	27 (54%)	7 (14%)	21 (36%)	8 (14%)	22 (37%)	8 (14%)	3.646 (p=0.302)
Pneumothorax	4 (9%)	25 (56%)	2 (4%)	14 (31%)	4 (10%)	21 (50%)	1 (2%)	16 (38%)	0.712 (p=0.87)
Perforation	^	^	^	^	^	^	^	^	^
Malignancy	29 (16%)	39 (22%)	28 (16%)	82 (46%)	2 (13%)	31 (17%)	31 (17%)	92 (52%)	2.113 (p=0.549)
Tuberculosis	33 (22%)	9 (6%)	31 (21%)	76 (51%)	24 (17%)	17 (12%)	32 (23%)	67 (48%)	4.189 (p=0.242)
Pulmonary Fibrosis	5 (11%)	2 (%%)	23 (52%)	14 (32%)	5 (12%)	4 (10%)	15 (37%)	17 (41%)	2.538 (p=0.468)
Other	25 (13%)	32 (16%)	4 (2%)	135 (69%)	21 (9%)	45 (19%)	11 (5%)	159 (67%)	4.1 (p=0.251)
TOTAL	330 (15%)	415 (19%)	706 (32%)	727 (34%)	331 (15%)	415 (19%)	699 (32%)	768 (34%)	0.882 (p=0.83)

Table 4.31 Influence of consultant radiologist and reporting radiographer chest X-ray reports on most likely diagnosis by final diagnosis
^=no cases; COPD = chronic obstructive pulmonary disease

Nearly one third of consultant radiologist (retained correct 330, 15%; new correct 415, 19%) and reporting radiographer (retained correct 331, 15%; new correct 415, 19%) chest X-ray reports had a positive influence on the most likely diagnosis of the clinicians; that is chest X-ray reports regardless of source (consultant radiologist or reporting radiographer) led to an accurate diagnosis for patients. One third of consultant radiologist (727, 33%) and reporting radiographer (768, 35%) reports appeared to have no influence on the clinicians' most likely diagnosis (retained incorrect diagnosis). One third appeared to have a negative influence, where the initial (pre-CXR) correct diagnosis was changed to an incorrect post-CXR diagnosis due to the consultant radiologist (706, 32%) or reporting radiographer (699, 32%) chest X-ray report. There was no significant difference in influence on clinicians' decisions between a consultant radiologist or reporting radiographer report (Chi-square χ=0.882, p=0.83).

The most positive influence of both consultant radiologist (25 cases, 56%) and reporting radiographer (21 cases, 50%) on clinicians' most likely diagnosis was for a pneumothorax. Reports which had the most detrimental influence of clinicians' diagnoses were cases which had no clinically significant disease (new incorrect most likely diagnosis) for both consultant radiologist (387 cases, 46%) and reporting radiographer (385 cases, 47%) chest X-ray reports. There was no apparent difference between the consultant radiologist and reporting radiographer chest X-ray reports for any of the final clinico-radiological diagnoses (Table 4.31).

4.4.5.5 Uncorrected diagnostic confidence in the most likely diagnoses

Of the 4,391 diagnostic confidences included in the analysis, there was a significant average overall increase in confidence of 8.5 (95%CI 7.8 – 9.2) from 71.8 to 80.3 (t=24.628, p<0.0001) for all chest X-ray reports. The changes in uncorrected diagnostic confidence between consultant radiologist and reporting radiographer chest X-ray reports are presented in Table 4.32 and Figure 4.8.

Reporting Practitioner	Uncorrected Most Likely Diagnostic Confidence	
	Pre-CXR (95% CI)	Post-CXR (95% CI)
Consultant Radiologist	71.0 (70.1 – 71.9)	80.4 (79.7 – 81.1)
Reporting Radiographer	72.5 (71.7 – 73.3)	80.2 (79.6 – 80.9)

Table 4.32 Consultant radiologist and reporting radiographer pre-CXR and post-CXR uncorrected most likely diagnostic confidence

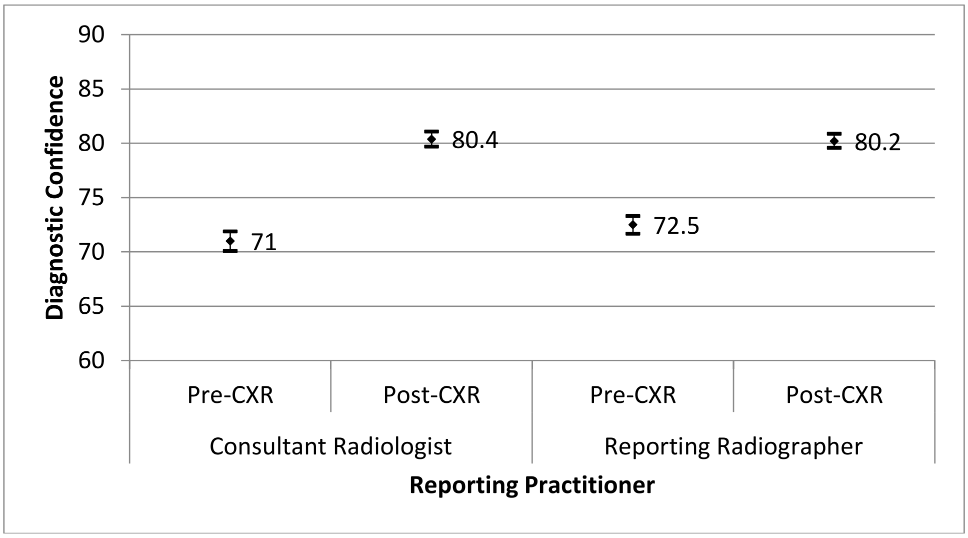


Figure 4.8 Consultant radiologist and reporting radiographer pre-CXR and post-CXR uncorrected most likely diagnostic confidence (with 95% CIs)

Chest X-ray reports from both consultant radiologist (71.0 to 80.4) and reporting radiographers (72.5 to 80.2) increased the uncorrected diagnostic confidence of clinicians for the most likely diagnosis. The clinician uncorrected most likely diagnostic confidence for reporting radiographer chest X-ray reports was non-inferior to the consultant radiologist chest X-ray reports ($t=23.81$, $p<0.0001$). Clinicians are more likely to instigate treatment decisions when they have a higher degree of confidence in their diagnosis, and these results suggests that chest X-ray reports from both consultant radiologists and reporting radiographers assist in this process.

4.4.5.6 Uncorrected diagnostic confidence in the most likely diagnoses for clinicians of different experience

The uncorrected confidence in the most likely diagnosis for consultant radiologist and reporting radiographer chest X-ray reports was examined for clinicians of different experience. The results are presented in Figure 4.9.

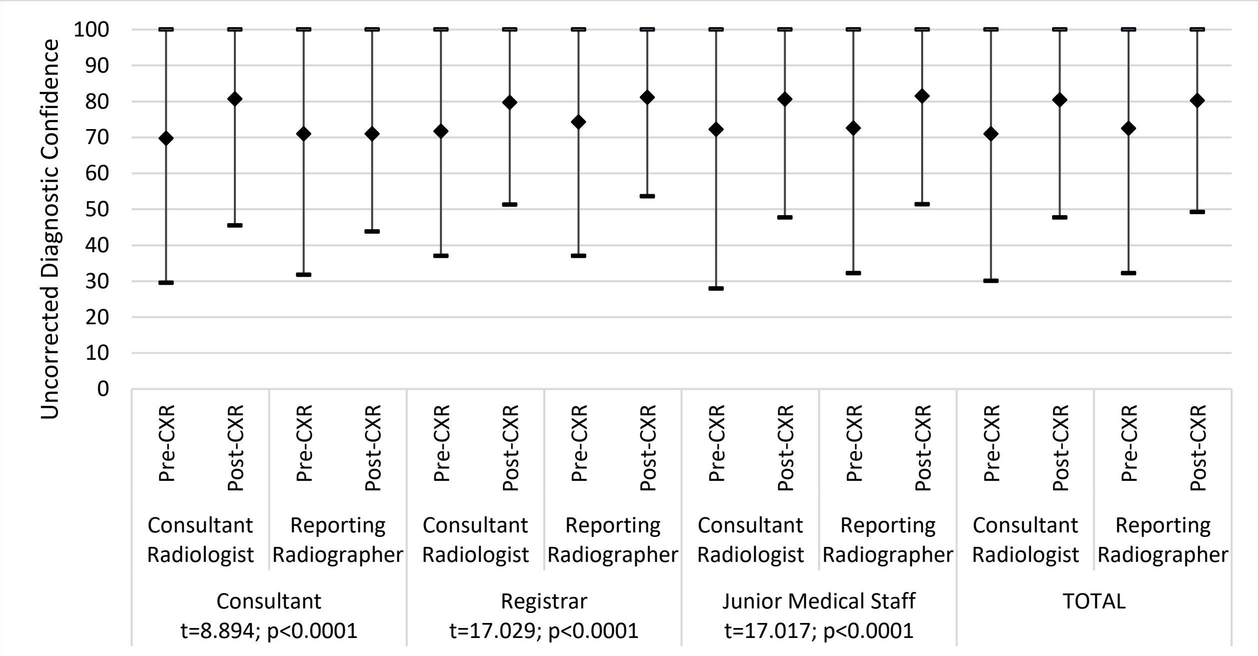


Figure 4.9 Uncorrected average most likely diagnostic confidence (with 95% CIs) for consultant radiologists and reporting radiographer chest X-ray reports for clinicians of different experience

Chest X-ray reports from both consultant radiologists and reporting radiographers increased the uncorrected diagnostic confidence of clinicians for the most likely diagnosis regardless of clinician experience. The clinician uncorrected most likely diagnostic confidence for reporting radiographer chest X-ray reports was non-inferior to the consultant radiologist chest X-ray reports (t=91.492, p<0.0001) for all observers and for clinicians of different experience.

4.4.5.7 Corrected diagnostic confidence in the most likely diagnoses

The most likely diagnoses for each case was compared to the final clinico-radiological diagnosis for accuracy and the corrected most likely diagnostic confidence calculated according to the Tsushima methodology (Chapter 3.3.6). Of the 4,391 diagnostic confidences included in the analysis, there was a significant average overall decrease of 67.0 (95%CI 64.1 – 69.8) from 71.8 to 4.8 (t=-46.61, p<0.0001) for all chest X-ray reports. This means that the chest X-ray reports of both consultant radiologists and reporting radiographers had a positive influence on the clinicians' diagnostic decision-making in only a small majority of cases, with little overall benefit when all cases were considered (0 = neutral/no benefit, -200 = very detrimental, 200 = very beneficial). The changes in corrected diagnostic confidence between consultant radiologist and reporting radiographer chest X-ray reports are presented in Table 4.33.

Reporting Practitioner	Corrected Most Likely Diagnostic Confidence Pre-CXR (95% CI)	 Post-CXR (95% CI)
Consultant Radiologist	71.0 (70.1 – 71.9)	5.02 (1.23 – 8.81)
Reporting Radiographer	72.5 (71.7 – 73.3)	4.61 (0.84 – 8.39)

Table 4.33 Consultant radiologist and reporting radiographer pre-CXR and post-CXR uncorrected most likely diagnostic confidence

There was little overall benefit in clinicians' diagnostic confidence in the most likely diagnosis when the post-CXR diagnosis was corrected for accuracy using the Tsushima method.[277]

Overall, the net contribution of both the consultant radiologist and reporting radiographer chest X-ray reports had little overall influence on the confidence of clinicians' in their most likely diagnosis once the most likely diagnosis was corrected for accuracy. Chest X-ray reports for both consultant radiologist (pre-CXR 71.0 to post-CXR 5.02) and reporting radiographers (pre-CXR 72.5 to post-CXR 4.61) decreased the corrected diagnostic confidence of clinicians for the most likely diagnosis. There was no apparent difference in the clinician corrected most likely diagnostic confidence for reporting radiographer chest X-ray reports compared to the consultant radiologist chest X-ray reports (t=0.048, p=0.093) but non-inferiority of the reporting radiographer reports was not established (null hypothesis was not rejected; p>0.05).

4.4.5.8 Corrected diagnostic confidence in the most likely diagnoses for clinicians of different experience

The corrected confidence in the most likely diagnosis for consultant radiologist and reporting radiographer chest X-ray reports was examined for clinicians of difference experience. The results are presented in Figure 4.10.

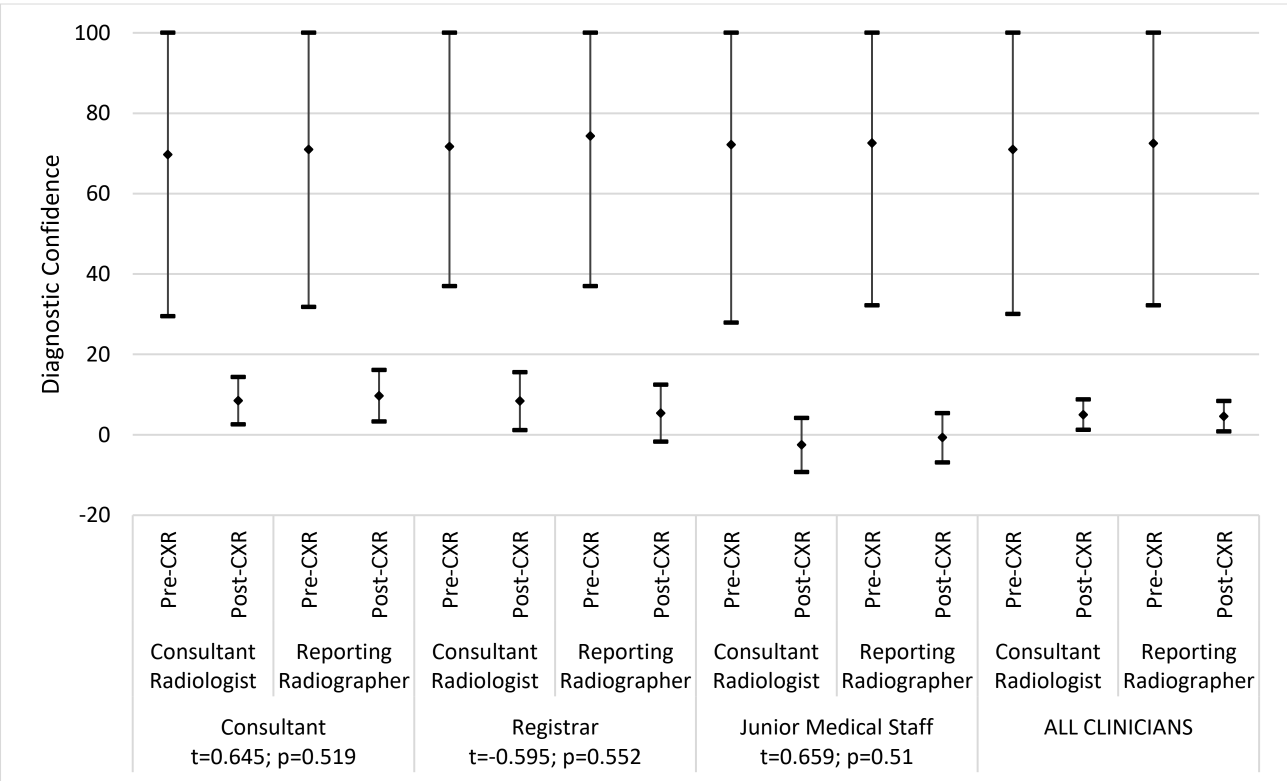


Figure 4.10 Average corrected most likely diagnostic confidence (with 95% CIs) for consultant radiologists and reporting radiographer chest X-ray reports for clinicians of different experience

The chest X-ray reports from both consultant radiologists and reporting radiographers decreased the corrected most likely diagnostic confidence regardless of clinician experience. Once corrected for accuracy of the diagnosis, the chest X-ray reports of both consultant radiologists and reporting radiographers had little overall influence on the diagnostic confidence regardless of clinician grade, with 95% confidence intervals close to, or crossing zero. There was no apparent difference between the influence of consultant radiologist and reporting radiographer chest X-ray reports on the corrected most likely diagnosis of all grades but non-inferiority of the reporting radiographer reports was not established as the null hypothesis was not rejected ($p>0.05$).

4.4.5.9 Corrected most likely diagnostic confidence for different diagnoses

The corrected clinician confidence in the most likely diagnosis for consultant radiologist and reporting radiographer chest X-ray reports was examined for the different final clinico-radiological diagnoses and is presented in Table 4.34.

Final Diagnosis (case note review)	Most Likely post-CXR corrected diagnostic confidence Consultant Radiologist (95%CI)	Most Likely post-CXR corrected diagnostic confidence Reporting Radiographer (95%CI)
No Significant Disease	11.1 (4.85 – 17.36)	4.2 (-2 – 10.4)
Infection	9.67 (1.0 – 18.3)	19.3 (10.3 – 28.25)
Cardiac/Pulmonary Oedema	-15.1 (-171.7 – 141.1)	-22.1 (-88.1 – 143.9)
Pleural Effusion	46.0 (-112 – 200)	37.6 (-119 – 193.9)
COPD	-24.7 (-200 – 156)	-37.5 (-200 – 179)
Pneumothorax	79.8 (-145.4 – 200)	68.2 (-150 – 200)
Perforation	^	^
Malignancy	7.0 (-182.4 – 196.4)	4.7 (-174 – 183.4)
Tuberculosis	-25.4 (-184.4 – 133.5)	-5.7 (-173.8 – 162.4)
Pulmonary Fibrosis	-4.5 (-118.1 – 109.1)	7.4 (-125.7 – 140.5)
Other	1.2 (-164.1 – 166.4)	11.2 (-176.4 – 185.6)
TOTAL	5.0 (-175.3 – 185.3)	4.6 (-176.4 – 185.6)*

Table 4.34 Influence of consultant radiologist and reporting radiographer chest X-ray reports on the corrected most likely diagnostic confidence by final diagnosis
COPD = chronic obstructive pulmonary disease; ^ = no cases; * p=0.208

The corrected most likely diagnostic confidence shows considerable variability, both in terms of the final clinico-radiological diagnosis but also within each diagnosis with very wide 95% confidence intervals. Both consultant radiologist and reporting radiographer chest X-ray reports had a positive influence in cases of pneumothorax (consultant radiologist 79.8, 95%CI -145.4 – 200; reporting radiographer 68.2, 95%CI -150 – 200) and pleural effusion (consultant radiologist 46, 95%CI -112 – 200; reporting radiographer 37.63, 95%CI -119 – 193.9). Chest X-ray reports for cases when the final diagnosis was chronic obstructive pulmonary disease (COPD) had a detrimental influence on clinicians diagnostic confidence, both for consultant radiologist (-24.7, 95%CI -200 – 156) and reporting radiographer (-37.5, 95%CI -200 – 179) reports. There was no apparent difference between the influence of consultant radiologist and reporting radiographer chest X-ray reports on clinicians' corrected most likely diagnostic confidence for different final diagnoses (F=1.344, p=0.208).

4.4.6 Clinician most serious diagnosis

The diagnosis that the clinician thinks is most likely will have the greatest influence on patient care and management decisions;[233] however there are often other diagnoses that are still considered within a differential diagnosis for a patient. Some of these diagnoses, while not the most likely, could potentially have a significant impact on patient outcome. In order to capture these diagnoses, and the potential influence that they may have on clinicians' diagnostic decision-making, the clinicians were asked to indicate a most serious diagnosis for all cases (Chapter 3.3.5). For the purposes of this study, the most serious diagnosis was "*the condition that the clinician would not want to miss in this patient, even if it is very unlikely"*(p.21).[233]

4.4.6.1 Most serious pre chest X-ray report diagnoses

The most serious diagnosis given by clinicians based solely on the clinical case summary (pre-CXR diagnosis) was compared to the final clinico-radiological diagnosis. The clinicians' most serious pre-CXR diagnoses are presented in Table 4.35.

Pre-CXR Most Serious Diagnosis	Consultants		Registrars		Junior Medical Staff		Total		
	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	All
No Significant Disease	20	0	17	0	0	0	37	0	**37**
Infection	331	136	187	42	269	47	787	225	**1012**
Cardiac/Pulmonary Oedema	34	71	125	98	136	74	295	243	**538**
Pleural Effusion	41	0	19	11	52	0	112	11	**123**
COPD	23	1	0	0	0	0	23	1	**24**
Pneumothorax	58	15	93	12	56	6	207	33	**240**
Perforation	177	0	70	0	109	0	356	0	**356**
Malignancy	265	87	189	57	274	86	728	230	**958**
Tuberculosis	165	89	101	23	164	39	430	151	**581**
Pulmonary Fibrosis	63	27	33	0	58	5	154	32	**186**
Other	74	21	121	52	139	33	334	106	**440**
TOTAL	1251	447	955	295	1257	290	3463	1032	**4495**

Table 4.35 Initial clinician most serious diagnoses based on clinical case summary (pre-CXR diagnosis)
COPD = chronic obstructive pulmonary disease

Infection (1,012 of 4,495; 23%) and malignancy (958 of 4,495; 21%) were the most frequent pre-CXR most serious diagnosis given by the clinicians, regardless of experience. The most serious diagnosis given by clinicians based solely on the case summaries (pre-CXR) was only correct in 23% of cases (1,032 of 4,495). Consultants (447 of 1,698; 26%) performed marginally better than registrars (295 of 1,250; 24%) and junior medical (290 of 1,547; 19%) clinicians.

4.4.6.2 Influence of chest X-ray reports on clinicians' most serious diagnoses compared with diagnoses reached without supporting chest X-ray report

After a minimum washout period of seven days, the clinicians were presented with the same clinical case summaries for each case and either a consultant radiologist or reporting radiographer chest X-ray report associated with the case. The post-CXR most serious diagnosis was compared to the initial most serious (pre-CXR) diagnosis provided by the clinician, and to the final clinico-radiological diagnosis for changes in diagnosis and for accuracy of the post-CXR diagnosis. Clinicians were blinded to the reporting practitioner who provided the report.

4.4.6.3 Change in clinician most serious diagnosis with chest X-ray reports

Cases in which the pre-CXR most serious diagnosis did not change with the chest X-ray report (retained diagnosis) and those that produced a change in clinician diagnosis (new diagnosis) are presented in Table 4.36.

Clinician Experience	Most Serious Diagnosis Retained	New
Consultant	739	925 (56%)
Registrar	570	655 (53%)
Junior Medical Staff	682	820 (55%)
Total	1991	2400 (55%)

Table 4.36 Chest X-ray reports which produced a change in clinician most serious (post-CXR) diagnosis

Approximately half of the chest X-ray reports produced a new clinician most serious diagnosis (2400 of 4391, 55%), with little variation according to clinician experience.

The proportions of consultant radiologist and reporting radiographer chest X-ray reports that produced a new clinician most serious diagnosis are presented in Table 4.37.

Clinician Experience	Most Serious Diagnosis Consultant Radiologist		Most Serious Diagnosis Reporting Radiographer	
	Retained	New	Retained	New
Consultant	403	498 (55%)	336	427 (56%)
Registrar	304	295 (49%)	266	360 (58%)
Junior Medical Staff	290	388 (57%)	392	432 (53%)
Total	997	1181 (54%)	994	1219 (55%)

Table 4.37 Cases which produced a new most serious diagnosis post-CXR

Approximately 50% of both consultant radiologist (n=1,181 of 2,178; 54%) and reporting radiographer (n=1,219 of 2,213; 55%) chest X-ray reports were associated with a change in the clinicians' most serious diagnosis, with similar proportions for clinicians of different experience.

4.4.6.4 Correct most serious diagnoses (retained and new)

The influence that chest X-ray reports had on clinicians' diagnostic decision-making was determined using the same methodology as for the most likely diagnosis (Chapter 3.3.5, Figure 4.6). The influence of the chest X-ray reports on the most serious diagnoses for clinicians with different experience are presented in Table 4.38.

Clinician Experience	Most Serious post-CXR diagnosis Consultant Radiologist				Most Serious post-CXR diagnosis Reporting Radiographer			
	Beneficial		Detrimental	Neutral	Beneficial		Detrimental	Neutral
	Retained Correct	New Correct	New Incorrect	Retained Incorrect	Retained Correct	New Correct	New Incorrect	Retained Incorrect
Consultant	84 (9%)	107 (12%)	153 (17%)	557 (62%)	82 (11%)	99 (13%)	118 (15%)	464 (61%)
Registrar	50 (8%)	78 (11%)	88 (15%)	383 (64%)	75 (12%)	68 (11%)	76 (12%)	407 (65%)
Junior Medical Staff	57 (8%)	75 (11%)	69 (10%)	477 (70%)	51 (6%)	98 (12%)	105 (13%)	570 (69%)
Total	191 (9%)	260 (12%)	310 (14%)	1417 (65%)	208 (9%)	265 (12%)	299 (14%)	1411 (65%)

Table 4.38 Influence of consultant radiologist and reporting radiographer chest X-ray report on the most serious diagnoses of clinicians of different experience

There were 451 (of 2,178, 21%) consultant radiologist and 473 (21%) of reporting radiographer chest X-ray reports which were associated with a positive influence on the clinicians' most serious diagnosis. No apparent difference between consultant radiologist and reporting radiographer chest X-ray reports was found (Chi-square χ=0.893, p=0.827) overall or when experience of clinician was examined (consultant χ=1.894, p=0.595; registrar χ=6.7, p=0.082; junior medical staff χ=4.955, p=0.175). The small proportion of cases in which the chest X-ray report had a positive influence on the most serious diagnosis (*the condition that the clinician would not want to miss in this patient, even if it is very unlikely*)(p.21)[233] is to be expected, as this diagnosis was structured to capture the infrequent but serious diagnoses that the clinician was considering for each patient.

4.4.6.5 Consultant radiologist and reporting radiographer chest X-ray reports influence on clinicians' most serious diagnosis by final clinico-radiological diagnosis

The influence that consultant radiologist and reporting radiographer chest X-ray reports had on the most serious diagnosis of clinicians' is shown in Table 4.39.

Final Diagnosis (case note review)	Most Serious post-CXR diagnosis Consultant Radiologist Retained Correct	New Correct	New Incorrect	Retained Incorrect	Most Serious post-CXR diagnosis Reporting Radiographer Retained Correct	New Correct	New Incorrect	Retained Incorrect	Chi-square (p value)
No Significant Disease	^	60 (7%)	^	776 (93%)	^	52 (6%)	^	761 (94%)	0.397 (p=0.529)
Infection	50 (12%)	61 (14%)	59 (14%)	254 (60%)	47 (11%)	66 (16%)	62 (15%)	243 (58%)	0.565 (p=0.904)
Cardiac/Pulmonary Oedema	61 (27%)	19 (8%)	52 (23%)	98 (43%)	76 (29%)	21 (8%)	53 (20%)	112 (43%)	0.606 (p=0.895)
Pleural Effusion	6 (23%)	1 (4%)	^	19 (73%)	5 (21%)	3 (13%)	^	16 (67%)	1.27 (p=0.53)
COPD	1 (2%)	8 (16%)	^	41 (82%)	^	14 (24%)	^	45 (76%)	2.094 (p=0.351)
Pneumothorax	9 (20%)	19 (42%)	6 (13%)	11 (24%)	16 (38%)	14 (33%)	2 (5%)	10 (24%)	4.667 (p=0.198)
Perforation	^	^	^	^	^	^	^	^	^
Malignancy	15 (8%)	32 (18%)	101 (57%)	30 (17%)	14 (8%)	42 (24%)	96 (54%)	26 (15%)	1.798 (p=0.615)
Tuberculosis	21 (14%)	16 (11%)	56 (38%)	56 (38%)	18 (13%)	17 (12%)	48 (34%)	57 (41%)	0.606 (p=0.895)
Pulmonary Fibrosis	10 (23%)	9 (20%)	8 (18%)	17 (39%)	5 (12%)	6 (15%)	7 (17%)	23 (56%)	3.313 (p=0.372)
Other	18 (9%)	35 (18%)	28 (14%)	115 (59%)	27 (11%)	30 (13%)	31 (13%)	148 (63%)	2.798 (p=0.424)
TOTAL	191 (9%)	260 (12%)	310 (14%)	1417 (65%)	208 (9%)	265 (12%)	299 (14%)	1441 (65%)	0.893 (p=0.827)

Table 4.39 Influence of consultant radiologist and reporting radiographer chest X-ray reports on clinicians' most serious diagnosis stratified by clinico-radiological diagnosis

The majority of consultant radiologist (1,417 of 2,178; 65%) and reporting radiographer (1,441 of 2,213; 65%) chest X-ray reports did not appear to influence clinicians' most serious diagnosis (retained incorrect diagnosis). The proportions of consultant radiologist and reporting radiographer chest X-ray reports that influenced the clinicians' most serious diagnosis in a positive way (new correct diagnosis or retained correct diagnosis) were similar. Consultant radiologist chest X-ray reports influenced the clinicians to retain a correct diagnosis in 191 cases (of 2,178, 9%) and reach a new correct diagnosis in 260 (of 2,178, 12%). This compared to the 208 (of 2,213, 9%) reporting radiographer chest X-ray reports which influenced the clinician to retain a correct diagnosis, and influenced the clinician to reach a new correct diagnosis for 265 cases (of 2,213, 12%). There was no apparent difference between the influence of consultant radiologist and reporting radiographer chest X-ray reports on the most serious diagnosis of the clinicians when the final clinico-radiological diagnosis was considered (Chi-square χ=0.893, p=0.827).

4.4.6.6 Uncorrected diagnostic confidence in the clinicians' most serious diagnoses

Changes in the clinicians' most serious diagnostic confidence were assessed using the same methodology as for the most likely diagnosis (Chapter 3.3.5; Chapter 4.4.5.5).

Of the 4,391 diagnostic confidences included in the analysis, there was a significant average overall increase of 6.8 (95%CI 5.8 – 7.8) from 33.8 to 40.6 (t=12.793, p<0.0001) for all chest X-ray reports. The changes in uncorrected diagnostic confidence between consultant radiologist and reporting radiographer chest X-ray reports are presented in Table 4.40.

Reporting Practitioner	Uncorrected Most Serious Diagnostic Confidence	
	Pre-CXR (95% CI)	Post-CXR (95% CI)
Consultant Radiologist	33.5 (32.3 – 34.7)	39.3 (37.8 – 40.7)
Reporting Radiographer	34.0 (32.8 – 35.2)	41.9 (40.5 – 43.3)

Table 4.40 Consultant radiologist and reporting radiographer pre-CXR and post-CXR uncorrected most serious diagnostic confidence

Chest X-ray reports for both consultant radiologists (33.5 to 39.9) and reporting radiographers (34.0 to 41.9) increased the uncorrected diagnostic confidence of clinicians for the most serious diagnosis. The clinician uncorrected most serious diagnostic confidence for reporting radiographer chest X-ray reports was non-inferior to the consultant radiologist chest X-ray reports ($t=9.022$, $p<0.0001$). The increase in clinicians' uncorrected confidence in the most serious diagnosis for both consultant radiologist and reporting radiographer chest X-ray reports suggest that the beneficial influence of chest X-rays as an investigation is not dependent on the practitioner providing the report.

4.4.6.7 Uncorrected diagnostic confidence in the most serious diagnoses for clinicians of different experience

The uncorrected confidence in the most serious diagnoses for consultant radiologist and reporting radiographer chest X-ray reports was examined for clinicians of difference experience. The results are presented in Figure 4.11.

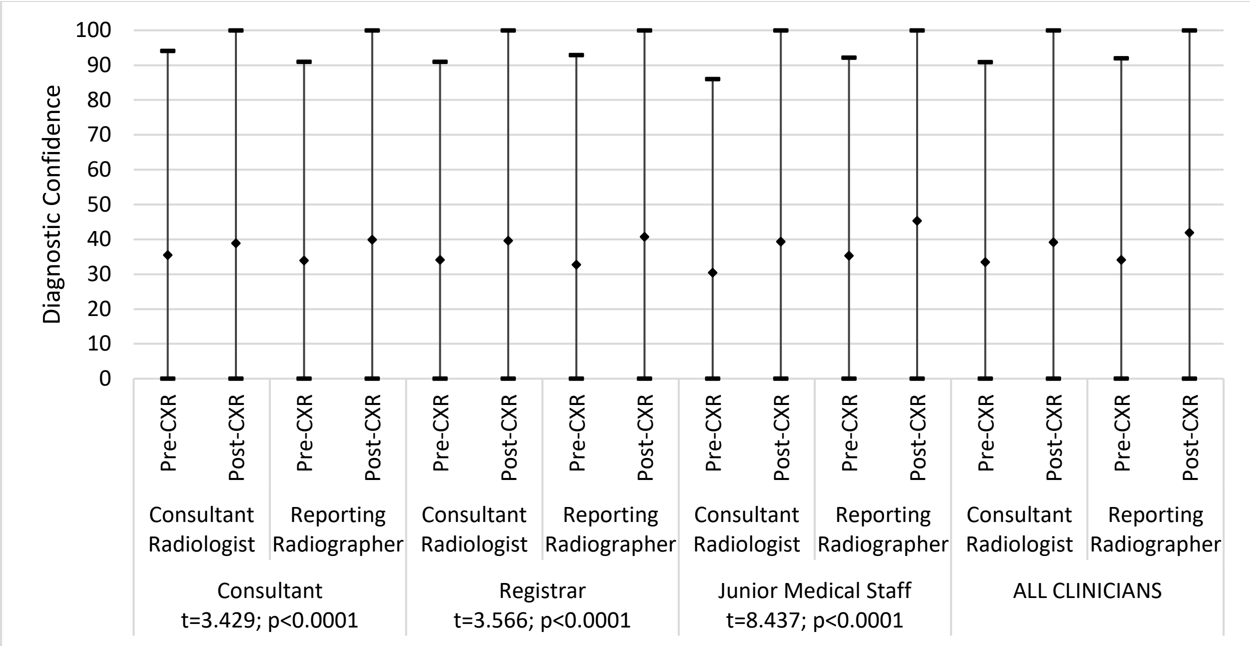


Figure 4.11 Uncorrected most serious diagnostic confidence (with 95% CIs) for consultant radiologist and reporting radiographer chest X-ray reports for clinicians of different experience

Chest X-ray reports from both consultant radiologists and reporting radiographers increased the uncorrected diagnostic confidence of clinicians for the most likely diagnosis although there was considerable variability with very wide 95% confidence intervals. Results were similar, regardless of clinician experience, thus both consultant radiologist and reporting radiographer reports had a positive overall influence on clinicians' diagnostic decision-making for the most serious diagnoses. The clinician uncorrected most likely diagnostic confidence for reporting radiographer chest X-ray reports was non-inferior ($p<0.05$) to the consultant radiologist chest X-ray reports for all clinicians regardless of experience. These results suggest that the beneficial influence that chest X-ray reports have on the confidence of clinicians in their most serious diagnoses is not influenced by the practitioner who provides the report.

4.4.6.8 Uncorrected diagnostic confidence in the most serious diagnoses for different diagnoses

The influence that consultant radiologist and reporting radiographer chest X-ray reports had on the uncorrected most serious clinician diagnostic confidence for different diagnoses are presented in Table 4.41.

Final Diagnosis (case note review)	Most Serious post-CXR uncorrected diagnostic confidence Consultant Radiologist (95%CI)	Most Serious post-CXR uncorrected diagnostic confidence Reporting Radiographer (95%CI)
No Significant Disease	28.1 (0 – 92.9)	31 (0 – 96.3)
Infection	39.6 (0 – 100)	42.4 (0 – 100)
Cardiac/Pulmonary Oedema	53 (0 – 100)	53.1 (0 – 100)
Pleural Effusion	48.6 (0 – 100)	57.3 (0 – 100)
COPD	49.3 (0 – 100)	54.7 (0 – 100)
Pneumothorax	63 (0 – 100)	56.1 (0 – 100)
Perforation	^	^
Malignancy	56.4 (0 – 100)	57.6 (0 – 100)
Tuberculosis	44.5 (0 – 100)	47.1 (0 – 100)
Pulmonary Fibrosis	48.5 (0 – 100)	43.1 (0 – 100)
Other	39.1 (0 – 100)	43.9 (0 – 100)
TOTAL	39.2 (0 – 100)	41.9 (0 – 100)

Table 4.41 Clinicians' uncorrected most serious diagnostic confidence for consultant radiologist and reporting radiographer chest X-ray reports for different final clinico-radiological diagnoses

The uncorrected clinician most serious diagnostic confidence was similar for consultant radiologist and reporting radiographer chest X-ray reports regardless of the final clinico-radiological diagnosis. There was considerable variation within each diagnosis, with wide 95% confidence intervals for both consultant radiologist and reporting radiographer chest X-ray reports. There was no apparent difference in the influence on uncorrected most serious diagnostic confidence for different diagnoses between consultant radiologist and reporting radiographer chest X-ray reports (F=0.562, p=0.829).

4.4.6.9 Corrected clinician most serious diagnostic confidence

The most serious diagnosis for each case was compared to the final clinico-radiological diagnosis for accuracy, and the corrected most serious diagnostic confidence calculated using the Tsushima method (Chapter 3.3.6).[277] Of the 4,391 diagnostic confidences included in the analysis, the chest X-ray reports produced a large average overall decrease in the confidence that clinicians' had in the most serious diagnoses of 31.0 (95%CI 29.1 – 32.9) from 33.8 to 2.8 (t=-31.001, p<0.0001) for all chest X-ray reports. This means that the chest X-ray reports only had a positive influence on the clinicians diagnostic decision-making in a small majority of cases, with little overall benefit (0 = neutral/no benefit, -200 = very detrimental influence, +200 = very beneficial). The changes in corrected clinician diagnostic confidence for the most serious diagnosis of consultant radiologist and reporting radiographer chest X-ray reports are presented in Table 4.42.

Reporting Practitioner	Corrected Most Serious Diagnostic Confidence Pre-CXR (95% CI)	Post-CXR (95% CI)
Consultant Radiologist	33.5 (0 – 90.9)	3.0 (0.56 – 5.41)
Reporting Radiographer	34.1 (0 – 92.0)	2.6 (0.15 – 5.05)

Table 4.42 Consultant radiologist and reporting radiographer pre-CXR and post-CXR corrected most serious diagnostic confidence

There was little overall benefit in clinicians' diagnostic confidence in the most serious diagnosis when confidence was corrected for the accuracy of the post-CXR diagnosis. Chest X-ray reports for both consultant radiologists (33.5 to 3.0) and reporting radiographers (34.1 to 2.6) decreased the corrected diagnostic confidence of clinicians for the most serious diagnosis. There was no apparent difference in the clinician corrected most serious diagnostic confidence for reporting radiographer chest X-ray reports compared to the consultant radiologist chest X-ray reports (t=0.048, p=0.093). Non-inferiority of the influence of the reporting radiographer chest X-ray reports was not established (t=-0.067, p=0.947).

4.4.6.10 Corrected diagnostic confidence in the most serious diagnoses for clinicians of different experience

The corrected confidence in the most serious diagnosis for consultant radiologist and reporting radiographer chest X-ray reports was examined for clinicians of difference experience. The results are presented in Table 4.43.

Clinician Experience	Corrected Most Serious Diagnostic Confidence			Non-Inferiority
	Reporting Practitioner	Pre-CXR (95%CI)	Post-CXR (95%CI)	
Consultant	*Consultant Radiologist*	35.5 (0 – 94.1)	6.4 (-117.8 – 130.5)	t = 0.166 p = 0.868
	Reporting Radiographer	33.9 (0 – 91.0)	6.1 (-120.1 – 132.4)	
Registrar	*Consultant Radiologist*	34.1 (0 – 91.0)	3.3 (-109.2 – 115.8)	t = -2.696 p = 0.007
	Reporting Radiographer	32.7 (0 – 92.9)	-3.4 (-120.8 – 114.1)	
Junior Medical Staff	*Consultant Radiologist*	30.4 (0 – 86.0)	-1.8 (-106.2 – 102.7)	t = 2.913 p = 0.004
	Reporting Radiographer	35.3 (0 – 92.2)	3.9 (-104.4 – 112.1)	
TOTAL	*Consultant Radiologist*	33.5 (0 – 90.9)	3 (-112.3 – 119.2)	
	Reporting Radiographer	34.1 (0 – 92.0)	2.6 (114.9 – 120.1)	

Table 4.43 Corrected most serious diagnostic confidence for consultant radiologists and reporting radiographer chest X-ray reports for clinicians of different experience

The experience of the clinician did appear to have had an influence on the corrected most serious diagnostic confidence (F=5.008, p=0.007). The chest X-ray reports of consultant radiologists (3.3 95%CI -109.2 – 115.8) had a more positive influence on the most serious diagnostic confidence of registrars when compared to reporting radiographer (-3.4; 95%CI 120.8 – 114.1) chest X-ray reports (t=-0.2696, p=0.007). Conversely, for junior medical staff, the reporting radiographer (3.9; 95%CI -104.4 – 112.1) chest X-ray reports had a more positive influence on the clinicians' corrected most serious diagnostic confidence when compared to consultant radiologist (-1.8; 95%CI -106.2 – 102.7) chest X-ray reports (t=2.913, p=0.004). The wide 95% confidence intervals and small net corrected most serious diagnostic confidence for both consultant radiologist and reporting radiographer chest X-ray reports for all clinician experiences suggest that the overall influence of all chest X-ray reports on the clinicians' corrected most serious diagnostic confidence was small.

4.4.6.11 Corrected diagnostic confidence in the most serious diagnoses for different diagnoses

The corrected clinician diagnostic confidence in the most serious diagnosis for consultant radiologist and reporting radiographer chest X-ray reports was examined for the different final diagnoses and is presented in Table 4.44.

Final Diagnosis (case note review)	Most Serious post-CXR corrected diagnostic confidence Consultant Radiologist (95%CI)	Most Serious post-CXR corrected diagnostic confidence Reporting Radiographer (95%CI)
No Significant Disease	8.8 (-77.7 – 95.2)	3.4 (-79.4 – 86.2)
Infection	7.4 (-117.9 – 132.8)	12.9 (-116.8 – 142.7)
Cardiac/Pulmonary Oedema	-24 (-147 – 98.9)	-27.5 (-54 – 99.1)
Pleural Effusion	-19.8 (-175.9 – 136.4)	-7 (-200 – 192.1)
COPD	8.3 (-120.7 – 137.3)	14.8 (-114.3 – 143.9)
Pneumothorax	39.8 (-138.2 – 200)	10.4 (-178.8 – 199.6)
Perforation	^	^
Malignancy	10.6 (-109.8 – 131.1)	21.6 (-109.6 – 152.7)
Tuberculosis	-21.6 (-154.3 – 111.2)	-14.4 (-144.7 – 115.8)
Pulmonary Fibrosis	-11.9 (-148.6 – 124.9)	-3.7 (-107.3 – 100)
Other	8.6 (-97.3 – 114.5)	8.4 (-103.6 – 120.3)
TOTAL	3 (-112.3 – 118.2)	2.6 (-114.9 – 120.1)

Table 4.44 Influence of consultant radiologist and reporting radiographer chest X-ray reports on the corrected most serious diagnostic confidence by final diagnosis
COPD = chronic obstructive pulmonary disease; ^ = no cases

The corrected most likely diagnostic confidence shows considerable variability, both between diagnoses but also within each diagnosis with very wide 95% confidence intervals. Both consultant radiologist and reporting radiographer chest X-ray reports had a positive influence in clinicians' confidence in the most serious diagnosis for cases of pneumothorax (consultant radiologist 39.8 95%CI -138.2 – 200; reporting radiographer 10.4 95%CI -178.8 – 199.6) and chronic obstructive pulmonary disease (consultant radiologist 8.3 95%CI-120.7 – 137.3; reporting radiographer -14.8 95%CI -114.3 – 143.9). There was no apparent difference between the influence of consultant radiologist and reporting radiographer chest X-ray reports on clinicians' corrected most likely diagnostic confidence for different final diagnoses (F=0.187, p=0.665).

4.4.7 Summary of performance for chest X-ray reports of consultant radiologists and reporting radiographers on clinicians' diagnostic decision-making

Clinicians' diagnostic decisions did not appear to be influenced by whether the report was provided by a consultant radiologist (CR) or reporting radiographer (RR). A similar proportion of consultant radiologist and reporting radiographer reports resulted in clinicians reaching a correct most likely or most serious diagnosis (CR=63%; RR=60%) and there was no apparent difference for different diagnoses or for clinicians of different experience.

Importantly both consultant radiologist and reporting radiographer chest X-ray reports had a beneficial influence on clinician uncorrected most likely (CR=80.4; RR=80.2) and most serious (CR=39.3; RR=41.9) diagnostic confidence, with the non-inferiority of the reporting radiographer chest X-ray reports established. When clinician diagnostic confidence was corrected for accuracy using the Tsushima methodology,[277] both the consultant radiologist and reporting radiographer chest X-ray reports demonstrated little overall influence for the most likely diagnosis (CR=5.02; RR=4.61) or most serious diagnosis (CR=3.0; RR=2.6). For clinicians' corrected most likely diagnostic confidence and corrected most serious diagnostic confidence there did not appear to be any difference between consultant radiologist and reporting radiographer reports, but non-inferiority was not established.

Chapter 5 – Discussion

5.0 Discussion

Prior to this study, no previous work has directly compared consultant radiologists and reporting radiographers in chest X-ray reporting. The non-inferiority of reporting radiographer diagnostic accuracy has been demonstrated across a range of methodologies; unweighted JAFROC, weighted JAFROC and inferred ROC. The diagnostic accuracy of the consultant radiologists and reporting radiographers in the current study is comparable to that of consultant radiologists in previous research.[77 82 228 251 281 282] In a further new finding, which group reports on chest X-rays (consultant radiologist or reporting radiographer) appears to have little or no influence on clinicians' diagnostic decision-making. These results suggest that reporting radiographers, having completed an appropriate accredited postgraduate programme of study, can safely complement consultant radiologists in terms of chest X-ray reporting in clinical practice.

The first part of this study, a robust assessment of the diagnostic accuracy of chest X-ray interpretation, confirms that the diagnostic accuracy of reporting radiographers when interpreting adult chest X-rays is non-inferior to that of consultant radiologists (Chapter 4.3). Sources of potential bias were identified and minimised for each element of the diagnostic accuracy study; image bank case selection, reference standard, observer selection and measurement of the results. A clinically representative image bank was constructed through use of a random stratified sample of chest X-rays performed over a consecutive 12 month period, and included a clinically representative spectrum of disease, a range of disease severity and conspicuity. The image bank was correlated with local audit and comparison made with national clinical datasets. Inclusion of cases from a range of referral sources ensured a diverse range of clinical demographics.[206] All practitioners who participated in the diagnostic accuracy study were qualified consultant radiologists or reporting radiographers who were currently interpreting chest X-rays in clinical practice, with a range of experience (consultant radiologists 3 – 31 years; reporting radiographers 1 – 9 years).[206]

The use of two independent, expert consultant chest radiologists produced a valid reference standard diagnosis for all cases included in the study, and was applied to all participant observer interpretations independently by two arbiters. The reference standard diagnosis was obtained independent to the reports of the participant reporting practitioners, and the same reference standard diagnosis was used for all cases. All image interpretation occurred independently and blinded to any other radiology report and with the relevant clinical history provided by the requesting clinician. With the exception of a small number of cases which were not interpreted by some practitioners due to oversight (consultant radiologists n=5, <1%; reporting radiographers n=8, <1%), all reporting practitioners interpreted an identical image bank. Two independent arbiters compared the reports to the reference standard diagnosis, blinded to the origin of the report (practitioner or profession). These measures have ensured that the results of the study are a reliable measure of diagnostic accuracy.

Part 2 of this study (Chapter 4.4) examined the influence that consultant radiologist and reporting radiographer chest X-ray reports had on clinicians' diagnostic decision-making using a rigorous study design. The results suggest that there is little difference between consultant radiologist and reporting radiographer chest X-ray reports. Non-inferiority of reporting radiographer chest X-ray reports was confirmed for some outcomes (uncorrected most likely diagnostic confidence and uncorrected most serious diagnostic confidence) for all clinicians regardless of experience. Non-inferiority was not established for all outcomes (corrected most likely diagnostic confidence and corrected most serious diagnostic confidence). The use of a range of cases, from different referral sources and with a wide spectrum of pathologies, ensured that the diagnostic confidence assessment was representative of a typical clinical caseload. Multiple clinician participants, with a range of experience, ensured that the influence of individual characteristics was minimised. Random, stratified allocation of consultant radiologist and reporting radiographer reports (Chapter 3.3.2) ensured that results were representative. The use of multiple reports of the same case from different reporting practitioners (consultant radiologists and reporting radiographers) reduced bias due to case selection. All clinician diagnostic decisions were made independently and blinded to the origin of the chest X-ray report (consultant radiologist or reporting radiographer). Correcting the post-chest X-ray diagnostic confidence to reflect the accuracy of the diagnosis reflected the clinical implications for instances where a clinician was 'confidently wrong' in their decision.[275] The 12-24 month delay between the chest X-ray and final clinico-radiological diagnosis minimised disease progression bias. Presentation of short case vignettes is an established methodology in medical education, as is the use of changes in pre and post investigation confidence in formulating diagnostic decisions, so did not introduce unfamiliar tasks to the clinician participants. The methodological rigour of the study has ensured that there is confidence that the results represent the influence that consultant radiologist and reporting radiographer chest X-ray reports have on clinicians' diagnostic decision-making.

As outlined in Chapter 2.7 marked variation between observers can exist when interpreting chest X-rays and that there is limited current evidence that examines the accuracy of reporting radiographer chest X-ray interpretation. The majority of previous studies have concentrated on a single task, for example nodule detection[180 227 228 283] or lung cancer screening,[33 34] unlike the current study that included a range of pathologies within the image bank. Other work has assessed the performance of radiographer abnormality detection[226] or preliminary clinical evaluation[284 285] rather than clinical reporting by qualified and practising reporting radiographers. The only other previous study that has examined the diagnostic accuracy of reporting radiographer chest X-ray reporting was in an academic setting,[41] but no direct comparison was made with the performance of consultant radiologists.

This current study is the first that has directly compared the diagnostic accuracy of a cohort of consultant radiologists and qualified reporting radiographers when interpreting a bank of adult chest X-rays selected from clinical practice and which includes a broad range of pathologies. The use of an image bank that closely reflects clinical practice enabled extrapolation of the findings of the current study to the expected performance of reporting radiographers in clinical practice.

The study has demonstrated statistically similar agreement in clinical practice between each expert and the consultant radiologists and reporting radiographers (Chapter 4.2). No apparent statistical difference was found on McNemar test; expert radiologist – consultant radiologist (p=0.701) and expert radiologist – reporting radiographer (p=0.629). The inter-observer agreement in the current study was lower than that reported by Woznitza *et al*. where agreement between a single reporting radiographer and three consultant radiologists was Kappa=0.83, Kappa=0.91 and Kappa=0.91.[42] The higher inter-observer agreement in the earlier study may have been due to incorporation and verification bias as the consultant radiologists compared the reporting radiographer interpretation to their own for accuracy.[206] In contrast, the current study was designed so the expert chest consultant radiologists performed their independent review blinded to the clinical report.

The influence that radiology reports have on clinicians' diagnostic decision-making is an important intermediate outcome measure in radiology research.[62 165 184] Diagnostic decision-making is the first part of the efficacy hierarchy which incorporates aspects outside of radiology as a surrogate measure for patient outcome.[57] The most recent evidence has concentrated on the influence that computed tomography [172 286] and magnetic resonance imaging[162 163] reports have on clinicians' diagnostic thinking. The work of Lusted in 1977 was the last major study that examined the influence of X-ray reports on clinicians' diagnostic decision-making.[233] In the current study, consultant radiologist and reporting radiographer chest X-ray reports resulted in comparable changes in clinicians' diagnostic decision-making. Similar numbers of consultant radiologist (n=1,368; 63%) and reporting radiographer (n=1,337; 60%) chest X-ray reports was associated with a correct clinician diagnosis (p=0.103), with little difference for clinicians of different experience (consultant, registrar, junior medical staff) and for different diagnoses. A total of 1,686 reports (consultant radiologist 810, 37%; reporting radiographer 876, 40%) where the clinician did not arrive at the correct diagnosis.

This may have a detrimental impact on patient care, however imaging investigations often only form one part of the diagnostic pathway, and so the negative influence may be moderated in clinical practice. It does however highlight the considerable cost burden, both in terms of patient outcome, radiation and finance that diagnostic errors can have.[287]

When the uncorrected diagnostic confidence was compared, using the methodology of Lusted,[233] the performance of the reporting radiographers (post-CXR uncorrected most likely diagnostic confidence 80.2 of 100) was non-inferior to that of consultant radiologists (post-CXR uncorrected most likely diagnostic confidence 80.4 of 100; p<0.0001). Lusted did not consider that the radiology report could have been incorrect or have resulted in the clinicians' considering an incorrect diagnosis,[233] a significant limitation at that time given the more recent evidence on diagnostic accuracy.[74 82 83 187 214] The work of Dixon and colleagues also did not consider inter-observer variation in diagnostic accuracy in their evaluations of the influence on magnetic resonance imaging reports on clinicians' diagnostic decision-making.[158 162-164]

Diagnostic confidence of the clinician post-CXR most likely diagnosis was corrected for the accuracy of the diagnosis using the Tsushima methodology.[277] Both the consultant radiologist (post-CXR corrected most likely diagnostic confidence 5.02 of 100) and reporting radiographer (post-CXR corrected most likely diagnostic confidence 4.61 of 100) chest X-ray reports in the current study showed very little overall benefit on clinicians' diagnostic decision-making. Although the study was designed to simulate clinical practice, the organic process of diagnostic decision-making often synthesises information from multiple sources simultaneously, including the response of the patient to treatment. This was not available given the structured nature of the assessment. This study is also the first to use the Tsushima methodology in a controlled environment, with previous work conducted in clinical practice.[275 277] These factors may have influenced the results of the current study. Another possible influence could have been the inclusion of multiple chest X-ray reports from different reporting practitioners for the same case. As demonstrated in the diagnostic accuracy study (Chapter 4.3), there was a range of diagnostic accuracy, with some cases producing incorrect diagnoses. These incorrect reports were included in the analysis, in an attempt to replicate the variation in accuracy of reports in clinical practice. The inclusion of multiple reports from multiple practitioners is a strength of the study design, as it allows the results to be generalised, but the net overall contribution of chest X-ray reports on clinicians' diagnostic decision-making may have been reduced as a result.

5.1 Methodological considerations and statistical analyses

There are several methodologies available to investigate diagnostic accuracy, including ROC, unweighted JAFROC and weighted JAFROC. As outlined in Chapter Appendix 12.4.1.2, JAFROC captures localisation and confidence rating for all abnormalities, rather than a global assessment of the entire case, and is a paradigm more closely aligned with clinical practice. The current study has determined the diagnostic accuracy of consultant radiologists and reporting radiographers using all three measures, to facilitate comparison with the existing literature. Few previous studies have compared the diagnostic accuracy of consultant radiologists for differences between ROC and JAFROC. Yamada and colleagues reported higher diagnostic accuracy for ROC analysis (chest X-ray mean AUC=0.77; tomosynthesis mean AUC=0.93) when compared to JAFROC (chest X-rays mean FoM=0.64; tomosynthesis mean FoM=0.88) but the authors did not discuss the reasons for such a discrepancy.[288] Kohli *et al.* also found the performance of both the chest consultant radiologists (n=6) and non-specialist radiologists (n=6) in nodule detection on a bank of 40 chest X-rays (20 abnormal) was higher for ROC (chest radiologists AUC=0.708; non-specialist radiologists AUC=0.659) than for JAFROC (chest radiologists FoM=0.487; non-specialists FoM=0.425).[282] These results are in line with the current study, with both the consultant radiologists (AUC=0.903; FoM=0.786) and the reporting radiographers (AUC=0.909; FoM=0.830) demonstrating a higher performance for ROC compared to JAFROC.

Kohli *et al*. also examined specific characteristics that can be associated with improved performance. They found that for JAFROC analysis the speciality of the radiologist, hours reading chest X-rays per week, satisfaction with chest X-ray interpretation and number of chest X-rays reported per year all correlated with higher accuracy.[282] In contrast, only hours per week reading chest X-rays was associated with improved performance for ROC analysis, and they postulate that the increased statistical power of JAFROC analysis facilitated the identification of these important factors that can influence accuracy.[282] This justifies the use of JAFROC as the primary measure of diagnostic accuracy in the current study. As a suggestion to improve practice, Kohli *et al.* propose that specialisation, time spent reporting chest X-rays per week and the number of chest X-rays interpreted annually are modifiable and controllable elements that could act as a surrogate for overall experience.[282] The findings of this study concur with this proposition. The consultant radiologists in this study had greater years of experience, the current volume of chest X-rays reporting in clinical practice was greater for the reporting radiographers, although no statistically significant difference was found in diagnostic accuracy when accounting for years of experience or current volume of examinations reported.

The use of a non-inferiority approach within medical research, and particularly radiology, is limited. As outlined in Appendix 12.6, failure to reject the null hypothesis (p>0.05) does not equate to equivalence/non-inferiority of the two investigations. Dedicated statistical methods are required to reduce bias and provide robust results. Despite recognised standards for non-inferiority studies,[48] a recent review of the radiology literature found poor compliance. Of the 38 studies published in two high impact radiology journals over a five and a half year period, only one study (3%) used an appropriate statistical method.[289] The majority of research (29 of 38, 76%) incorrectly reported non-inferiority/equivalence based on a p value greater than 0.05 for a standard superiority approach.

Many of the identified studies (28 of 38, 74%) failed to report a sample size calculation. Thus, a false claim of non-inferiority could have been due to an underpowered study. The current study adhered to current best practice guidance; a pre-defined clinically insignificant margin was used, sample size calculations were performed and appropriate statistical analysis conducted.[48] The results of the current study, which found reporting radiographers non-inferior to consultant radiologists in chest X-ray interpretation, are based on an appropriate and robust methodological approach. As less invasive treatment options, lower cost or lower radiation dose examinations are developed non-inferiority studies are expected to grow. The approach used in the current study could act as a benchmark for comparative analysis of consultant radiologist and reporting radiographer diagnostic accuracy.

5.2 Agreement of expert chest consultant radiologists in establishing the reference standard diagnosis

The significant body of evidence that examines the performance of observers when interpreting chest X-rays confirms that considerable inter-observer variation exists. Accordingly, this was taken into account for data analysis in the current study. Other factors, that include the nature of the reporting task, characteristics of the observers, whether the study was conducted in clinical practice or a controlled environment and technological developments also need to be considered when comparing current results to the literature.

The current study found the expert chest radiologists agreed in 52 (of 180 cases, 29%) of the normal cases and 79 (of 180 cases, 44%) of abnormal cases, with moderate agreement (Kappa Ҟ=0.48; 95%CI 0.36 – 0.59; p<0.0001) for a random, stratified sample of chest X-rays with a range of pathologies. These results are comparable to those found in the historical literature that scrutinises the tuberculosis and lung cancer screening programmes. Overall agreement in the diagnosis of tuberculosis ranged from 66% to over 99%,[65 66 70 194] while agreement on abnormalities identified in lung cancer screening has been reported as between 10%[73] and 67%.[71] Some of this variation has been due to the different radiographic techniques used, with many screening programmes replacing conventional 35 cm x 43 cm X-ray film with miniature 35 mm film.[65-67] Advances in technology also limit comparison with the current study. Chest X-rays are now acquired using computed radiography (CR) or digital radiography (DR), although more recent work (1980-2003) that used improved film-screen radiography and full size 35 x 43 cm X-ray film still report a significant error rate.[73 77 78 211] The inter-expert chest radiologist agreement in the current study compared favourably to the contemporary work of Singh *et al*. who reported fair agreement (K=0.34) in the detection of lung nodules on chest X-rays by expert thoracic radiologists,[79] and the agreement (K=0.38) reported by Balabanova *et al*. when identifying tuberculosis.[75]

Both the work of Robinson *et al*. who found three consultant radiologists agreed in only 61 of 100 chest X-rays (K=0.63 – 0.68)[188] and Tudor *et al*. when a bank of 50 X-rays were interpreted with clinical history (K=0.58)[214] reported slightly higher agreement than the inter-expert radiologist agreement in the current study (K=0.48). The influence that disease prevalence has on radiologists' diagnostic accuracy is equivocal, with conflicting results.[255 279 290 291] Previous work by Reed *et al.* suggested that prevalence expectation did not alter expert radiologist accuracy in lung nodule detection.[255] This is unlike the earlier work of Egglin and Feinstein,[290] and the subsequent work of Reed *et al.*[291] and Littlefair and colleagues,[279] who all found differences in radiologist accuracy with both actual higher disease prevalence and higher expected prevalence. Controlled studies of diagnostic performance often use enhanced proportions of abnormal cases to test observers for rare pathologies or to reduce sample sizes, but this may not be representative of clinical practice.[254] What is not in doubt is the influence of disease prevalence (the proportion of abnormal cases in an image bank) on the Kappa statistic, with higher prevalence introducing bias.[292 293] This may partly explain the higher inter-observer agreement reported by others.[188 214] The agreement between expert chest consultant radiologists in the current study suggests that they are comparable with the performance of experts found in the literature.[77 82 188 214]

The historical and contemporary literature that has examined chest X-ray accuracy of consultant radiologists is consistent in demonstrating variation in performance. This is persistent despite advances in technologies used to acquire the chest X-rays (miniature film, 35 x 43 cm films, computed radiography and digital radiography). Variation occurs irrespective of the image interpretation task; lung nodule detection, diagnosis of tuberculosis and when reporting a spectrum of pathologies. The inclusion of cases in the image bank only where both expert chest radiologists agreed has accounted for this variation and so justified the choice of reference standard diagnosis for the diagnostic accuracy study.

Analysis at a case level of the diagnostic accuracy of the cohorts of consultant radiologists and reporting radiographers confirmed that there were very few cases in which all participants were correct, both normal and abnormal. The broad spectrum of pathologies, in proportions that mirror clinical practice based on local (Appendix 3) and national audit,[245] have ensured that the case mix for the image bank used in the current is reflective of clinical practice. Thus, the diagnostic accuracy results produced from this robust and reliable image bank, allow conclusions on reporting practitioner diagnostic accuracy to be drawn.

5.3 Expert chest radiologist agreement with consultant radiologist and reporting radiographer chest X-ray reports from clinical practice

As detailed in Chapter 3.2.2, to mimic clinical practice cases were drawn from a randomised, stratified sample of adult chest X-rays performed for clinical reasons. To minimise bias, the expert chest radiologists performed their interpretations blinded to the clinical report.[206] This facilitated the comparison of expert chest consultant radiologists (CC1 and CC2) to the clinical reports provided by both consultant radiologists (CR) and reporting radiographers (RR). This study has found comparable levels of normal – abnormal agreement between clinical reports provided by consultant radiologists and reporting radiographers (detailed in Chapter 4.2) though these were lower than that reported by Woznitza *et al.*,[42] perhaps owing to incorporation and verification bias[206] in the latter study.

Comparable report concordance was also found between the expert chest consultant radiologists and the clinical reports of consultant radiologists. However, studies that have examined performance in clinical practice have reported mixed results. The review by Quekel and colleagues,[77] who utilised expert chest radiologists to perform retrospective review of new lung cancer diagnoses, reported slightly higher accuracy (81%) when compared to the current study. The work of Austin *et al.* also examined cases of missed lung cancer and reported similar findings.[294] Twenty-seven cases missed in clinical practice by 18 radiologists were included in the analysis, and 22 chest X-rays were retrospectively reviewed by six expert chest radiologists. Accuracy of the expert radiologists was 74%, more in line with the findings of this study. Of importance, *Austin et al.* reported that despite bias from the targeted nature of the expert review (missed lung cancer), the lung cancer was missed by at least one expert radiologist in 16 (73%) of cases.[294] Due to the targeted nature of these reviews, no normal cases were included and the work focused on a single pathology rather than a range of diseases. This could partially explain the results. The radiologist and radiographer performance in the current study was comparable (agreement 75%) to that of Herman *et al*., and the use of 100 randomly selected cases in their study more closely mimics the methodology employed in the current study.[264]

There is a general paucity of evidence which has examined the performance of radiographers interpreting chest X-rays in clinical practice. Prior to the current study, there have only been four studies that have examined radiographer chest X-ray reporting. Two of these historical studies were conducted as part of lung cancer screening research in the 1970s. The analysis of Flehinger *et al.* reported encouraging results of radiographer first reading of screening chest X-rays, with error rates of 2% and 3.2% for the two radiographers.[34] This may not have been an accurate reflection of performance, as cases that were deemed normal by both radiographer and radiologists were not subjected to expert radiologist review (verification bias), and the radiographers performed their interpretation in conjunction with the radiologist (incorporation bias).[206] More recently, Sonnex *et al*. assessed the performance of radiographers in a specialist thoracic centre, and found that they correctly identified normal (specificity 99%) and abnormal chest X-rays (sensitivity 90%).[226] A crucial distinction, which must be made when comparing the current study to this previous work, is that the radiographers in the previous studies were not reporting radiographers.[33 34 226] They were required to distinguish between normal and abnormal cases only, and did not provide a clinical report or a diagnosis. The reported accuracy of radiographers in the previous studies may have been elevated.[206] The results of the current study, which used 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. The only previous work that utilised trained reporting radiographers reporting chest X-rays in clinical practice was performed by Robinson.[37] A small number of chest X-rays from the emergency department were included as part of a larger study, and agreement with a radiologist report found in 52 normal and 31 abnormal cases (of 112, 74%).[37]

5.4 Diagnostic accuracy of consultant radiologists and reporting radiographers when interpreting a bank of adult chest X-rays

There is a considerable body of evidence which examines the diagnostic accuracy of chest X-ray interpretation in a controlled (image bank) setting. The majority of observers in these studies are consultant radiologists, with their performance compared to new technologies including digital subtraction[295 296] and tomosynthesis.[231 288] Traditional comparison between consultant radiologists with a variety of non-radiology medical practitioners of varying experience have also been performed.[75 230 297] Accuracy, in terms of true negative and true positive, is the most common summary measure of performance, especially for studies that compare consultant radiologists to non-radiology medical practitioners. Receiver operator characteristic (ROC) methodology is also frequently used.[75 82 263 298] There is a growing body of work that has employed the free response methodology and in particular the jack-knife alternate free response received operator characteristic curve methodology.[228 231 279 281 284 299-301] Little work has been done that has examined the diagnostic accuracy of reporting radiographer chest X-ray interpretation and these few studies have examined nodule detection rather than interpreting a broad spectrum of pathologies.[81 227 228]

5.4.1 Diagnostic accuracy studies of reporting radiographer chest X-ray interpretation

A small number of previous studies have investigated the diagnostic accuracy of reporting radiographer chest X-ray interpretation. Piper *et al*. examined the performance of a cohort of 40 reporting radiographers (n=4,000 chest X-rays) at the end of an accredited postgraduate training programme.[41] They found high sensitivity and specificity, 95.4% and 95.9% respectively, for an image bank with a diverse range of pathologies. The average sensitivity and specificity of both consultant radiologists (sensitivity 69.7%, specificity 80.9%) and reporting radiographers (sensitivity 78.1%, specificity 85.2%) in the current study was lower. The current results are not directly comparable due to methodological differences; sensitivity and specificity were calculated differently[279] as sensitivity and specificity were not the primary outcome measures. Robinson conducted the only other study that examined the diagnostic accuracy of qualified reporting radiographers in chest X-ray interpretation.[37] Robinson's study was a feasibility study examining radiographer reporting of 112 emergency department chest and abdominal X-ray examinations. He used two radiographers only and did not differentiate between the chest and abdominal examinations. He found an overall accuracy of 74% (52 concordant normal reports, and 31 concordant abnormal reports) when compared to a single consultant radiologist report. Unlike the current study, neither Piper *et al*.[41] nor Robinson[37] compared the performance of reporting radiographers to consultant radiologists.

The reporting radiographer sensitivity (78.1%) and specificity (85.2%) in the current study is in line with that reported in the recent work of Ekpo *et al*. who found moderate accuracy in a cohort of Nigerian radiographers, with 80% sensitivity and specificity when interpreting a bank of chest X-rays in a controlled setting.[285] The current results are also comparable to that of Sheft *et al*.[33] They found similar performance in the ability of two radiographers to identify cases suspicious for cancer on chest X-rays (after training radiographer false positive 13% and 19%; false negative 8% and 4%) to that of five trainee radiologists (mean false positive 9% and false negative 8%) and three consultant radiologists (mean false positive 7% and false negative 6%).[33] As outlined above, direct comparison with the current study cannot be made due to the nature of the task (clinical report vs. abnormality detection) and the educational experience of the participants (accredited postgraduate training vs. limited formal training).

5.4.2 Diagnostic accuracy of chest X-ray interpretation which have used jack-knife alternate free response receiver operator characteristic curve (JAFROC) methodology

The consultant radiologists (FoM=0.786) and reporting radiographers (FoM=0.830) in the current study compare favourably with the reported figures of merit (range 0.4 – 0.82) within the literature. A summary of all studies that have used JAFROC methodology to assess the diagnostic accuracy of chest X-ray interpretation are included as a table in Appendix 2. Even for the range of study designs used, the current results are comparable across the spectrum and suggest that the cohort of reporting radiographers in the current study report adult chest X-rays with similar performance to consultant radiologists demonstrated in previous research.

Only two previous studies have included a range of pathology in the image bank. Szucs-Farkas *et al*. reported figures of merit between 0.446 – 0.595 for consultant radiologists and a radiology registrar in an image bank that included lung cancer, lymphoma, sarcoidosis and Wegner's vasculitis.[253] Kasai *et al*. analysed accuracy of lung nodule (FoM=0.622) and thoracic vertebral fracture (FoM=0.585) detection.[302] The performance of both the consultant radiologists (FoM=0.786) and reporting radiographers (FoM=0.830) found in the current study compare favourably to this previous work.

Nodule detection studies, the bulk of previous work that has used JAFROC for diagnostic accuracy studies of chest X-ray interpretation, focus on a single clinical task rather than the spectrum of pathologies encountered in clinical practice. This does facilitate comparison with other variables, such as the use of new techniques including computer assisted diagnosis and tomosynthesis. To facilitate comparison with the literature with the current study, the figures of merit from the control reading, that is under normal reporting conditions without a new technique for example, have been used.

The studies that have examined the role that tomosynthesis has in the accuracy of lung nodule detection, Vikgren *et al*.,[301] Yamada *et al*.[288] and Doo *et al*.,[80] found average figures of merit of 0.40, 0.64 and 0.41 respectively for the control reading (chest X-rays without tomosynthesis). The performance of both the consultant radiologists (FoM=0.786) and reporting radiographers (FoM=0.830) in the current study compare favourably. However, these studies had small participant numbers (3-4) limiting generalisability, whereas the current study had a greater number of observers that assists when extrapolating the diagnostic accuracy to the wider population of reporting radiographers.

Both Brennen *et al*.[281] and Kohli *et al*.[282] compared the performance of subspecialist chest radiologists and general radiologists in nodule detection. The average performance of the general radiologists (FoM=0.68 without noise, FoM=0.69 with noise) and chest radiologists (FoM=0.65 without noise, FoM=0.68 with noise) found by Brennan *et al*.[281] was lower than the results of the current study (CR FoM=0.786, RR FoM=0.830), and superior to that of Kohli *et al*. (FoM=0.425).[282] A smaller number of cases was used by both Brennen *et al*. (n=30)[281] and Kohli *et al*. (n=40)[282] when compared to the current study (n=106). Fewer observers were also used by Kohli *et al*. (n=6 general consultant radiologists)[282] with Brennan *et al*. using a similar number of observers (n=15 general consultant radiologists, chest radiologists n=11)[281] to the current study (n=10 consultant radiologists).

Observer expectation and the influence that this may have on diagnostic accuracy has been examined in several ways using different variables. The work of Littlefair *et al*. used different anticipated levels of disease prevalence in their analysis,[279] McEntee and Quinn did not provide clinical history[303] and Robinson *et al*. used focused abnormality detection.[84] The controlled reading figures of merit for these studies, FoM=0.60,[279] FoM=0.55[303] and FoM=0.671[84] respectively, again are similar to the results of the current study. A smaller number of cases were included in all of these works, approximately 40 chest X-rays, when compared to the current study (n=106). A greater number of cases, with a broad spectrum of pathologies, enables the results of the current study to be generalised to a wide population of patients, rather than the small subset of consultant radiologist observers included in previous work.

The previous work that has used the jack-knife alternate free response methodology for the assessment of reporting radiographers in chest X-ray interpretation has concentrated on a single task, namely nodule detection, rather than a range of pathologies commonly encountered in clinical practice. Manning *et al*. found that the five radiographers after postgraduate training (FoM=0.82) performed marginally better that the eight consultant radiologists (FoM=0.80),[227] which is similar to the results of the current study (consultant radiologist FoM=0.786; reporting radiographer FoM=0.830). The number of radiologist (n=8) observers and the number of cases (n=120) are also similar to the current study, although fewer radiographer (n=5) observers were used, further strengthening the comparisons. Donovan and Litchfield included two reporting radiographers in their study.[228] Only the summary figure of merit was reported (FoM=0.72) for the expert group and they did not differentiate reporting radiographers (n=2) from consultant radiologists (n=8).[228] This was possibly due to the small numbers of observers. Another explanation for this grouping of consultant radiologists and reporting radiographers for analysis may have been that previous work conducted by the researchers had suggested comparable accuracy.[227] The work of Buissink *et al*., that reported a figure of merit of 0.677,[284] is not as comparable to the current study as the radiographer cohort had not completed accredited postgraduate training.

5.4.3 Diagnostic accuracy studies of chest X-ray interpretation which have used receiver operator characteristic curve methodology

Inferred receiver operator characteristic (ROC) curves were also calculated as part of the current study to facilitate comparison with the literature. A summary of the literature that utilised ROC curves for diagnostic accuracy of chest X-ray interpretation is presented in Appendix 1. Although both JAFROC and ROC are measures of diagnostic accuracy, the methodology varies and thus results are not directly comparable between the two designs. The calculation of inferred ROC facilitates the comparison of the diagnostic accuracy of the consultant radiologists and reporting radiographers in the current study with results of previous research.

As was found with many of the JAFROC studies, observer performance which used the ROC methodology was often comparing accuracy with an intervention, such as the inclusion of clinical history, dual energy imaging or for different medical specialities. The area under the ROC curve for the control reading have been used when comparing previous results to the current study. This further comparison with the literature adds additional strength to any conclusions that can be drawn on reporting radiographer chest X-ray interpretation. The majority of studies that have used receiver operator characteristic (ROC) curve methodology have been performed with consultant radiologist observers with considerable variation in sample size, both in terms of observers (range 4 – 162) and cases (range 30 – 247). Many of the studies used a balanced disease prevalence ratio (normal:abnormal 1:1), as was used in the current study, but few studies used a range of pathologies. The performance of both the consultant radiologists (average AUC=0.903) and reporting radiographers (average AUC=0.909) in this study compare favourably to those studies that included a range of pathologies. Good *et al*. [304] reported AUC between 0.78 – 0.98, the consultant radiologists in Potchen *et al*. [82] had an average AUC of 0.86 and Balabanova *et al*.[75] an average AUC of 0.81. When the current study is compared to previous work which used a range of pathologies but included chest X-rays in an image bank with other examinations, Berbaum *et al*.,[305] Tudor and Finlay[207] and Eng *et al*.[263] all report lower AUC for consultant radiologists than both the reporting radiographer and consultant radiologist participants in current study, 0.745, 0.88 and 0.81 respectively.

Work that examined observer detection of a single pathology, in general, reported lower AUC values than the current study. For nodule detection, Kashani *et al*.[295] and Shang *et al*.[252] found AUC values of 0.789 and 0.66, broadly comparable to Furhman *et al*.[306] who examined rib fracture detection (AUC = 0.73). These three studies compensated for a small number of observers, between four and eight, with a greater number of cases (100 – 129). A large number of cases may have been used in order to enable the results to be generalisable to a broad patient population but this is limited by the narrow focus of the interpretation task. The current study does not have these limitations, with a range of pathologies included within the 106 cases and a sufficient number of observers (CR n=10, RR n=11), yet has superior AUC values for both the consultant radiologists (AUC=0.903) and the reporting radiographers (AUC=0.909).

Berbaum and colleagues have investigated satisfaction of search, a source of diagnostic error where an abnormality may be missed due to the presence of additional abnormalities.[307-309] The two studies that used the same bank of 57 chest X-rays that included a range of pathologies evaluated the influence of a checklist and computer assisted diagnosis. They reported similar diagnostic accuracy of the consultant radiologists in the control reading (no checklist, no computer assisted diagnosis) in both studies, AUC=0.67[307]and AUC=0.681,[308] lower than the accuracy of the consultant radiologists (AUC=0.903) and reporting radiographers (AUC=0.909) in the current study. The accuracy of the consultant radiologists in a study that examined satisfaction of search using simulated lung nodules found was higher than the previous work, AUC=0.74,[310] although still lower that the current study. Comparison with the existing literature suggests that, for ROC analysis, the consultant radiologists in the current study are comparable to their peers, and therefore the performance of the reporting radiographers are also comparable to the wider population of consultant radiologists.

5.5 Influence of reporting radiographer chest X-ray reports on clinicians' diagnostic decision-making

The work of Lusted, a large multisite evaluation of the contribution that radiology reports have on clinicians' diagnostic thinking, represents the largest and most comprehensive analysis of the diagnostic impact of X-rays.[233] Of the 2,627 chest X-rays included in the study, the radiologist report produced a new most likely diagnosis in 13% (342 of 2,627), in contrast to the 51.2% (1,088 of 2,178) consultant radiologist and 48.8% (1,137 of 2,213) reporting radiographer chest X-ray reports in the current study. The proportion of new most likely diagnoses in the literature demonstrates similar variability. Work which has investigated cross sectional imaging reports new most likely diagnoses between 30% (37 of 125)[277] and 40% (71 of 118)[172] of cases for abdominal-pelvic CT and 27% (69 of 269)[164] and 56% (55 of 98)[163] for MRI reports of the knee and wrist. This variability may be due to the patient population included in the various studies; emergency department,[233] acute surgical,[172 277] and orthopaedic outpatient,[162-164] with the higher proportion of new most likely diagnoses in the current study partially due to the broad range of referral sources (emergency department, inpatient, outpatient). The results of the current study are at the upper end of the range of diagnostic accuracy, and does suggest that the findings in the current study are relevant to clinical practice.

The chest X-ray reports of all reporting practitioners in the current study increased the uncorrected most likely diagnostic confidence of the clinicians, from 71.0% to 80.4% and 72.5% - 80.2% for the consultant radiologist and reporting radiographer chest X-ray reports respectively, with no statistically significant difference ($p<0.0001$). Lusted also found an average increase in clinicians' most likely diagnostic confidence, from 9.1% to 59.9%.[233] The net increase of Lusted (50.8%) was higher than the current study (9.4% consultant radiologist, 7.7% reporting radiographer) but the initial clinician confidence was much lower, 9.4% in contrast to 71.0% and 72.5%.[233] The reason for such a large discrepancy in the initial most likely diagnostic confidence between the clinicians in the current study and the previous work of Lusted could be due to the setting of the task, controlled academic setting compared to clinical practice, and this may have changed clinician behaviour. Despite the lower overall increase in most likely diagnostic confidence in this current study, there was no statistically significant difference between the chest X-ray reports of the consultant radiologists and the reporting radiographers.

In contrast to the current study that stratified cases by a broad spectrum of pathologies, only cases that had infection or no clinically significant disease were subject to sub-analysis by Lusted.[233] Cases in which the clinician suspected a diagnosis of infection in Lusted's work accounted for the bulk of chest X-rays, 1,213 of 2,627 (46%), lower than the 24% (1,066 of 4,495) in the current study. A slightly lower proportion of cases in the current study (319 of 4,391; 7%) had a clinically confirmed infection (pneumonia or tuberculosis), compared to 10% (265 of 2,627) in Lusted's study. An important consideration, as discussed below, was that Lusted's diagnosis was based solely on the radiological report and did not incorporate clinical follow up.[233] Confirmed cases of infection in the current study were established at clinico-radiological review, and thus there is greater confidence that these cases truly represent infection.[233] Of the cases included in the current study, 36 (34%) were from patients who had no significant disease as the final clinico-radiological diagnosis, similar to the 24% (630 of 2,627) in Lusted.[233] This suggests that the use of emergency department referrals by Lusted, in contrast to the range of referral sources in the current study, may have increased the overall disease prevalence.

A large number (n=381) of emergency clinicians participated in Lusted's[233] evaluation compared to the current study (n=18) but the proportion of chest X-ray reports reviewed by clinicians with different experience were similar. Of the cases included in Lusted's analysis, consultants reviewed 879 (34%), registrars 823 (31%) and junior medical staff n=925 (35%)[233] compared to the current study, consultants n=1664 (38%), registrars n=1225 (29%) and junior medical staff n=1502 (34%). The current study found little difference in the proportion of consultant radiologist and reporting radiographer chest X-ray reports that produced a new diagnosis for clinicians of different experience, in line with Lusted.[233] As there was a comparable proportion of chest X-ray reports reviewed by clinicians of different experience in both studies, it could be expected that this similarity in influence on diagnostic confidence would occur in clinical practice.

The use of a single referral source (emergency department) and the evolution of diagnostic technology over the past 40 years limits the generalisability of Lusted's findings to contemporary practice and for direct comparison to the current study. The most significant limitation in Lusted's study was the interpretation of a single radiologist as the determinant of diagnosis.[233] The accuracy of the radiologist was unquestioned, as "selected chart review has very rarely shown contradiction of this diagnosis"(p.175).[145] This is in contrast to both the historical [67 194 265 268] and current [74 82 240 269] literature regarding observer variation in image interpretation.

In order to address this limitation, the Tsushima methodology was used to correct clinician confidence according to the accuracy of the post-chest X-ray diagnosis in the current study. This method has only previously been used in studies which investigated CT,[277 311] and the current study is the first to employ this methodology for chest X-rays. When confidence was corrected for the accuracy of the diagnosis, there was a significant reduction from 80.4% to 5.02% for consultant radiologist and 80.2% to 4.61% for reporting radiographer chest X-ray reports. This is in contrast to the previous work that has employed this methodology. Tsushima *et al*. found a positive influence of the radiology report in 118 of 125 CT scans and an average increase in diagnostic confidence of

40.5% (95%CI 30.1% - 69.2%) from 32.9% to 73.4%.[277] These results mirror the work of Ng and Palmer, who found a correct diagnosis on 50 (of 62 patients, 81%) and reported an average positive improvement in confidence of 37.7% (95% CI 20.1% – 55.4%) using the Tsushima method in a comparative analysis using different methodologies for calculating diagnostic confidence.[311] Both of these studies analysed considerably smaller numbers of radiology reports; 125[277] and 62,[311] in comparison to the 4,391 reports in the current study. The discrepancy between the current study, and previous work which reported a positive impact of radiology reports, may have been driven by the nature of the investigation (abdominal-pelvic CT compared to chest X-ray), the sample size used and the setting in which the study was conducted (clinical practice[277] compared to an controlled setting [current study]).

The task given to the participant clinicians in this study, arriving at a diagnosis based on a short case summary/vignette, is an established assessment technique in medical education.[312] In clinical practice, reports are synthesised organically with ongoing information that the clinicians receive, including further diagnostic tests, patient improvement or deterioration and repeated clinical assessment/review.[280] Tsushima *et al*. also only assessed clinicians' most likely diagnostic confidence,[277] rather than both the most likely and most serious diagnostic confidence as assessed in the current study and by Lusted.[233] The use of both in the current study may have influenced results. Future work, conducted in clinical practice, may help to clarify this further. The results of the current study, which found no difference in clinicians' diagnostic confidence between consultant radiologist or reporting radiographer chest X-ray reports, suggests that undertaking further work in the clinical setting could be performed without a negative influence on patient management decisions or outcomes.

All consultant radiologist and reporting radiographer chest X-ray reports from Part 1 (diagnostic accuracy study) were included in Part 2 (diagnostic decision-making). Reports that were known to be both correct and incorrect were used. The inclusion of incorrect chest X-ray reports may have had a detrimental impact on the clinicians' diagnoses and diagnostic confidence. If the incorrect consultant radiologist and reporting radiographer chest X-ray reports had been excluded from analysis this may have increased both uncorrected and corrected diagnoses and diagnostic confidence. The exclusion of incorrect chest X-ray reports would have limited external validity, as not all reports provided in clinical practice are correct. Clinicians may use correct and incorrect chest X-ray reports differently, especially when the radiological diagnosis deviates considerably from the clinical presentation and the results of other investigations. Further research in this area would be useful.

The current study, which utilised multiple reporting practitioners (consultant radiologists and reporting radiographers) and multiple clinicians, is the first to account for inter-practitioner variability when determining the influence of radiology reports on clinician diagnostic confidence. All previous work that has examined the influence of radiology reports on clinicians' diagnostic confidence has used a single report and single clinician review.[56 143 158 162-164 171 233 276] Even in studies that have corrected the clinicians' diagnostic decision for accuracy of the final (follow up) diagnosis,[172 275 277 286] only a single radiology report and single reviewing clinician decision has been evaluated. As outlined previously, the use of a single consultant radiologist report as the 'final' diagnosis fails to consider adequately the known error and variation in radiology interpretation. Similarly, the use of a single clinician's review of the radiology report does not account for differences in experience, knowledge, and personal characteristics of the clinician.[280] The use of multiple clinician review of a chest X-ray report, a strength of the current study design, may be a contributing factor to the little overall benefit in diagnostic confidence for both consultant radiologist and reporting radiographer reports in the current study. While there was little overall benefit for the corrected clinician diagnostic confidence, there was little difference between the reports of consultant radiologists and reporting radiographers.

5.6 Study limitations

The main limitation, for both the diagnostic accuracy and diagnostic decision-making studies, was that the work occurred in a controlled setting rather than clinical practice. The results of this study are promising, but as identified by several authors [63 183 254], this performance may not be transferrable to clinical practice. This approach did provide the advantage of increasing the number of participants, and all reporting practitioners interpreting the same cases with a robust reference standard diagnosis.

Although similar results were found between consultant radiologists and reporting radiographers, the contribution of chest X-ray reports on clinicians' diagnostic decision-making showed little overall benefit when the diagnosis was corrected for accuracy against the final clinico-radiological diagnosis obtained at case note review. Several variables, such as the use of multiple clinician review of the same case, review including the same chest X-ray report, and reports provided by different reporting practitioners could have been involved. Further investigation, including conducting an evaluation in clinical practice rather than a controlled setting, would provide further clarity. It would be feasible to conduct an exploratory study, perhaps focused in an acute setting such as the emergency department and medical assessment, to investigate further.

The reference standard diagnosis used in the diagnostic accuracy study was the consensus findings of two independent expert chest radiologists who had access to follow up imaging. Histology is frequently used as the reference standard diagnosis in other chest X-ray diagnostic accuracy studies, and it may be possible that some of the malignant cases could have been due to another pathology. This is mitigated by the use of CT (where available) as this is more frequently being used as the reference standard diagnosis in diagnostic accuracy studies, and the use of histology would have raised ethical concerns for the inclusion of other pathologies in the image bank.

Referrals from general practice were excluded from the current study due to logistical limitations in retrieving patient case notes for case summary construction and clinico-radiological diagnosis. This may limit the generalisability of the findings to referrals from outside an acute setting, but this is balanced by the inclusion of outpatient referrals. As such, it is thought that the results of the current study can be extrapolated to this patient group, but future work including these patients would be helpful.

Some of the limitations of the current study may be addressed with further research (Chapter 5.7).

5.7 Directions for future work

The current study has established that consultant radiologists and reporting radiographers interpret adult chest X-rays with comparable accuracy, and that there is little difference in the influence that these chest X-ray reports have on clinicians' diagnostic thinking in a controlled setting. This study, which builds on previous work,[41] is an essential first step and confirms that those reporting radiographers currently in clinical practice do not appear to compromise patient care or safety. The increased use of reporting radiographers to interpret chest X-rays in clinical practice is both an effective and efficient method for radiology departments to meet predicted activity increases and the requirement for new models of care.[9 12 101 313-317] This is supported by the results of the current diagnostic accuracy study (Chapter 4.3) and the comparison of expert radiologist agreement with clinical reports (Chapter 4.2) and previously published clinical audit data.[42]

In line with previous work, variation was found between reporting practitioners when interpreting the chest X-rays in the current study. This variation was found between consultant radiologists and reporting radiographers when reporting the image bank (Chapter 4.3), and also between expert radiologists when forming the reference standard diagnosis (Chapter 4.1). The cases that produced discordant expert radiologist interpretations were excluded from further analysis. Future work could include analysis of these cases to examine the source of discrepancy and comparison with other observers interpreting these cases such as radiologists and non-radiologist physicians.

Using the principles of evidence based medicine and evidence based practice,[51 59 104] the best current evidence supports chest X-ray interpretation by reporting radiographers. Diagnostic performance, not just of radiographer reporting but of all diagnostic investigations, can be different when taken from a controlled setting. Care must be taken when transferring the results from controlled assessments of diagnostic accuracy in to clinical practice.[45 62 254] When establishing radiographer chest X-ray reporting in clinical practice, routine audit data that is required for all radiology departments[318] should be published. This does not replace the need for a robust, high quality assessment of radiographer chest X-ray reporting, comparable to the work that examined skeletal reporting by radiographers.[22 186 234]

The study that examined the influence of consultant radiologist and reporting radiographer chest X-ray reports found promising results and this suggests that radiographer reporting of chest X-rays does not have an adverse influence on clinicians' diagnostic decision-making in a controlled setting. One possible reason for the limited influence of the chest X-ray report on diagnostic confidence corrected for accuracy is that the X-ray report was taken in isolation, rather than as part of a dynamic diagnostic work up which occurs in everyday practice. Future work could include a targeted intervention in an acute or emergency setting, where the influence of the chest X-ray reports produced by reporting radiographers could be evaluated as part of the diagnostic decision-making and management pathway of clinicians. This clinically based research would facilitate direct quantification of the role of the chest X-ray report.

The contribution of allied health practitioners to effective and efficient care often goes unnoticed, with limited evidence of impact.[319]. This may be in part due to the relative paucity of research evidence that examines the cost effectiveness and impact of radiographer roles, particularly advanced practice. The health economic assessment that has been performed did establish that immediate radiographer reporting of skeletal X-rays in an emergency setting is cost effective and provides better outcomes for patients.[141] Further work that explores the impact and cost effectiveness of radiographer chest X-ray reporting, for example on the lung cancer pathway, would be a valuable contribution.

5.8 Dissemination plans

Feedback to participants will be offered by email. For the diagnostic accuracy study participants will be provided with their accuracy results and the average performance of both professional groups (consultant radiologists and reporting radiographers). The clinicians who participated in the diagnostic decision-making study, will be offered the number of correct post-CXR diagnoses they made, and the average performance of the three clinician experience cohorts (consultant, registrars, junior medical).

A strategy for publication of the findings requires sequential writing of a paper to support release of the main findings. Information will be shared with the Society and College of Radiographers, with a report prepared outlining a summary of the methodology and results in line with the requirements agreed when funding was received.

5.9 Recommendations for practice

Clinical imaging is a fundamental component of many patient pathways. Radiology activity has seen sustained increases over the past decade.[94] The focus on improved patient outcomes and the challenges that an aging population with complex health needs is expected to bring, demand is showing no signs of abating.[95 320 321] The relative similarity of the performance of the reporting radiographers in the current study to the existing knowledge base provides further evidence that reporting radiographers can contribute to chest X-ray reporting in clinical practice.

There are significant, well documented, challenges facing radiology services across England with many departments struggling to meet existing demand.[3 13 15 20] The new streamlined and rapid approach to cancer diagnosis that is being advocated to improve patient outcomes [8 101] will only contribute to increased radiology workload. Integrated radiographer reporting has been shown as an effective way of ensuring patients receive high quality care,[17] and that the results of the current study suggest that increased radiographer reporting of chest X-rays could be used as part of the response to these demands.

It is crucial that clinical imaging departments adapt in order to improve, while a safe and high quality service is maintained.[43] Integrated radiographer reporting is a reasonable approach to tackle these issues and has been demonstrated to be a safe option for services. Patient safety could even be improved with radiographer chest X-ray reporting, for example by helping to avoid ‘never events’ related to nasogastric tubes by providing a contribution to the timely clinical reporting of chest X-rays.[322] The additional diagnostic capacity created by radiographer reporting of chest X-rays can also help meet legislative requirements for clinical reporting.[167 168]

In order to improve patient outcomes, there has been a move across the health service to concentrate care in specialised centres with high volumes.[323 324] Significant variation is found in chest X-ray interpretation accuracy, in historical and contemporary literature and the current study. Radiology is not unique, and there is a body of evidence that demonstrates that specialist radiologists interpret examinations with higher accuracy than non-specialist radiologists.[82 325-327] Expert specialist radiologists, although not immune to discordant interpretations, perform better than non-specialist radiologists do in chest X-ray reporting accuracy. There has been a shift within the UK, and many departments now have a significant proportion of X-ray examinations reported by trained reporting radiographers.[17 18] Number of cases reported per year, and time spent per week have been shown to improve accuracy.[282] There may come a time that the most appropriate person to report chest X-rays are specialist chest radiologists and specialist chest X-ray reporting radiographers.

Another option to improve reporting accuracy is double reporting, similar to the system employed by the NHS breast cancer screening programme for mammograms.[209] The high volume of chest X-rays performed every year,[94] coupled with significant reporting backlogs[14 15 328] and insufficient diagnostic capacity[8 12] may act as a barrier to the routine double reporting of chest X-rays.

Set on a background of reporting backlogs, many departments are increasingly using outsourced radiology services in an attempt to manage demand.[4 15 20] While allowing flexibility, this approach does not address the capacity issues driving unreported examinations and does not facilitate service redesign for optimum outcomes. A more sustainable method to meet current and anticipated increases in activity is to increase the contribution of radiographer reporting. Once trained and integrated into departments, chest X-ray reporting radiographers as part of a multidisciplinary team can increase reporting capacity without compromising in accuracy.

Diagnostic workforce shortages, including radiographers and consultant radiologists, makes it essential that the full skills and scope of practice for all practitioners needs to be maximised. Implementation of the full four tier structure,[102] which includes assistant, advanced and consultant practitioners, will be required if fundamental changes are to be made to patient pathways and service delivery.[9 317] One model could be to integrate assistant radiographic practitioners and advanced practitioner reporting radiographers for image acquisition and immediate reporting, thus maximising the benefit to the patient of the diagnostic procedure through the timely availability of high quality images and together with accurate reports. The results of this research confirm that this is a safe and achievable option for chest X-ray reporting and could facilitate implementation of the national optimal lung cancer pathway that advocates immediate reporting of chest X-rays referred from general practice.[329]

Efficiency within healthcare has always been important, more so recently with the unprecedented financial climate.[96] Radiographer reporting of chest X-rays has the potential to address some of these efficiency requirements[313 315 320 330] while maintaining an effective service[317 331] and providing the opportunity to develop and deliver new models of care. The findings of the current study indicate reporting radiographers are a step along that route knowing their reporting performance is comparable with consultant radiologists.

5.10 Conclusions

In a new contribution to the literature, this study shows that reporting radiographers and radiologists demonstrate similar levels of diagnostic reporting accuracy for chest X-ray interpretation. The performance of the reporting radiographers in the current study also compares well to previous work which determined the diagnostic accuracy of consultant radiologists, and is further evidence that reporting radiographers can report chest X-rays with accuracy comparable to consultant radiologists. That both the number of cases and number of observers in the current study is comparable to the literature is a strength of the current study, and suggests that the results are generalisable to a wider population of trained reporting radiographers.

Who reports on the X-ray, whether reporting radiographer or consultant radiologist, appears to have no influence on clinicians' decision-making. This is a further new finding that has important implications as this suggests that radiographer reporting of chest X-rays can be safely implemented into clinical practice, with no detrimental impact on clinicians' diagnostic decision-making.

References

1. Cochrane A. L. *Effectiveness and Efficiency: random reflections on health services*: The Nuffield Provincial Hospitals Trusts, 1972.
2. Royal College of Radiologists. Clinical Radiology UK Workforce Report 2011. 2012 London: Royal College of Radiologists
3. Royal College of Radiologists. Patients waiting too long for test results. London: Royal College of Radiologists, 2014.
4. Royal College of Radiologists. Clinical radiology UK workforce census 2015 report. 2016 London: Royal College of Radiologists
5. Canadian Institute for Health Information. Health Care in Canada: A Focus on Wait Times. 2012 Ottowa, Canada: Canaadian Institute for Health Information
6. Queensland Department of Health. Inquiry into Gold Coast X-ray reporting. Queensland, Australia, 2014.
7. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2015. 2015 Toronto, Canada: Canadian Cancer Society
8. Taskforce Independent Cancer. Achieving world-class cancer outcomes: A strategy for England 2015-2020. 2015 London
9. NHS England. Five Year Forward View. 2014 London: NHS England
10. NHS England. Achieving World-Class Cancer Outcomes: Taking the strategy forward. 2016 London: NHS England
11. NHS England. Seven Day Services Clinical Standards. 2016: NHS England
12. 2020 Delivery. Horizon Scanning: An evaluation of imaging capacity across the NHS in England. 2015: Cancer Research UK
13. Royal College of Radiologists. Unreported X-rays, computed tomography (CT) and magnetic resonance imaging (MRI) scans: Results of a snapshot survey of English National Health Service (NHS) trusts. London: Royal College of Radiologists, 2015.
14. Royal College of Radiologists. Unreported X-rays, computed tomography (CT) and magnetic resonance imaging (MRI) examinations: results of the September 2015 snapshot survey of English NHS acute trusts. 2015 London: Royal College of Radiologists
15. Royal College of Radiologists. Our patients are still waiting …. 2016 London: Royal College of Radiologists
16. Ratcliff C., Woznitza N. Re-audit case study: Optimising patient care in radiology through team-working. 2015 London: College of Radiographers
17. Woznitza N., Piper K., Rowe S. *et al.* Optimizing patient care in radiology through team-working: A case study. Radiography 2014;**20**(3):258-263.
18. Snaith B., Milner R., Harris M. A. Beyond image interpretation: Capturing the impact of radiographer advanced practice through activity diaries. *Radiography* 2016;**22**(4):e233-e238.
19. National Health Service and Community Care Act (c.19). United Kingdom, 1990.
20. Royal College of Radiologists. Clinical Radiology UK Workforce Census Report 2012. 2014 London: Royal College of Radiologists
21. Society and College of Radiographers. Scope of radiographic practice survey 2012. 2012 London: Society and College of Radiographers
22. Piper K., Paterson A., Ryan C. The implementation of a radiographic reporting service for trauma examinations of the skeletal system in 4 NHS trusts. 1999: NHS Executive South Thames
23. 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.
24. 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.
25. Sumkin J. H., Klaman H. M., Graham M. *et al.* Prescreening mammography by technologists: a preliminary assessment. *Am J Roentgenol* 2003;**180**(1):253-256.

26. Wivell G., Denton E. R., Eve C. B. *et al.* Can radiographers read screening mammograms? *Clin Radiol* 2003;**58**(1):63-67.
27. Tanaka R., Takamori M., Uchiyama Y. *et al.* Using breast radiographers' reports as a second opinion for radiologists' readings of microcalcifications in digital mammography. *Br J Radiol* 2015;**88**(1047):20140565.
28. Nightingale J., Hogg P. The role of the GI radiographer: a United Kingdom perspective. *Radiol Technol* 2007;**78**(4):284-290.
29. Meertens R., Brealey S., Nightingale J. *et al.* Diagnostic accuracy of radiographer reporting of computed tomography colonography examinations: a systematic review. *Clin Radiol* 2013;**68**(4):e177-190.
30. Piper K., Buscall K. L., Thomas N. MRI reporting by radiographers: Findings of an accredited postgraduate programme. *Radiography* 2010;**16**(2):136-142.
31. Brealey S., Piper K., King D. *et al.* Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist. *Eur J Radiol* 2013;**82**(10):e597-605.
32. Snaith B., Hardy M., Lewis E. F. Radiographer reporting in the UK: A longitudinal analysis. *Radiography* 2015;**21**(2):119-123.
33. 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.
34. 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. *Am J Roentgenol* 1978;**131**(4):593-597.
35. 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.
36. Loughran C. F. Reporting of fracture radiographs by radiographers: the impact of a training programme. Br J Radiol 1994;**67**(802):945-950.
37. Robinson P. J. Plain film reporting by radiographers--a feasibility study. *Br J Radiol* 1996;**69**(828):1171-1174.
38. Robinson P. J., Culpan G., Wiggins M. Interpretation of selected accident and emergency radiographic examinations by radiographers: a review of 11000 cases. *Br J Radiol* 1999;**72**(858):546-551.
39. Hardy M., Snaith B. The impact of radiographer immediate reporting on patient outcomes and service delivery within the emergency department: Designing a randomised controlled trial. *Radiography* 2011;**17**(4):275-279.
40. Snaith B., Hardy M. The perceived impact of an emergency department immediate reporting service: An exploratory survey. *Radiography* 2013;**19**(2):92-96.
41. Piper K., Cox S., Paterson A. *et al.* Chest reporting by radiographers: Findings of an accredited postgraduate programme. *Radiography* 2014;**20**(2):94-99.
42. 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.
43. Care & Quality Commission. Essential standards of quality and safety: Care & Quality Commission, 2010.
44. Royal College of Radiologists, Society and College of Radiographers. Team working in clinical imaging. 2012 London: Royal College of Radiologists and the Society and College of Radiographers
45. Knottnerus J. A., Buntinx F., van Weel C. General Introduction: evaluation of diagnostic procedures. In: Knottnerus J. A., Buntinx F., eds. *The evidence base of clinical diagnosis: theory and methods of diagnostic research*. 2nd Ed ed. Singapore: Blackwell Publishing, 2009.

46. Bossuyt P. M., Reitsma J. B., Bruns D. E. *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Br Med J* 2003;**326**(7379):41-44.
47. Bossuyt P. M., Reitsma J. B., Bruns D. E. *et al.* STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology* 2015;**277**(3):151516.
48. Piaggio G., Elbourne D. R., Pocock S. J. *et al.* Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *J Am Med Assoc* 2012;**308**(24):2594-2604.
49. Vandenbroucke J. P., von Elm E., Altman D. G. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Ann Intern Med* 2007;**147**(8):W163-W194.
50. Guyatt G. Evidence-based medicine. *ACP J Club (Ann Intern Med)* 1991;**14(suppl 2)**:A-16.
51. Sackett D. L., Rosenberg W. M. C., Gray J. A. M. *et al.* Evidence based medicine: what it is and what it isn't. *Br Med J* 1996;**312**(7023):71-72.
52. Nursing & Midwifery Council. The code: Standards of conduct, performance and ethics for nurses and midwives. London: Nursing & Midwifery Council, 2008.
53. Health and Care Professions Council. Standards of proficiency: Radiographers. 2013 London: Health and Care Professions Council
54. Greenhalgh T., Howick J., Maskrey N. *et al.* Evidence based medicine: a movement in crisis? *Br Med J* 2014;**348**:g3725.
55. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;**44**(3):Suppl:166-206.
56. Fineberg H. V., Bauman R., Sosman M. Computerized cranial tomography. Effect on diagnostic and therapeutic plans. *J Am Med Assoc* 1977;**238**(3):224-227.
57. Fryback D. G., Thornbury J. R. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**(2):88-94.
58. Hunink M. G. M., Krestin G. P. Study Design for Concurrent Development, Assessment, and Implementation of New Diagnostic Imaging Technology. *Radiology* 2002;**222**(3):604-614.
59. Hollingworth W., Jarvik J. G. Technology assessment in radiology: putting the evidence in evidence-based radiology. *Radiology* 2007;**244**(1):31-38.
60. National Institute for Health and Clinical Excellence. Medical Technologies Evaluation Programme Methods guide. 2011 London: National Institute for Health and Clinical Excellence
61. Mackenzie R., Dixon A. K. Measuring the effects of imaging: an evaluative framework. *Clin Radiol* 1995;**50**(8):513-518.
62. Brealey S. Measuring the effects of image interpretation: an evaluative framework. *Clin Radiol* 2001;**56**(5):341-347.
63. Brealey S., Scally A. J. Methodological approaches to evaluating the practice of radiographers' interpretation of images: A review. *Radiography* 2008;**14**(S1):e46-e54.
64. Chakraborty D. P. A brief history of free-response receiver operating characteristic paradigm data analysis. *Acad Radiol* 2013;**20**(7):915-919.
65. de Lorimer A. A. Mass roentgenography of the chest for the United States Army. *Radiology* 1942;**38**(4):462-472.
66. Hodges F. J. Fluorographic examination of the chest as a routine hospital procedure. *Radiology* 1942;**38**:453-461.
67. Levitin J. Ten-thousand chest examinations with the stereoscopic photoroentgen unit. *Am J Roentgenol* 1943;**49**(4):469-475.
68. Birkelo C. C., Chamberlain W. E., Phelps P. S. *et al.* Tuberculosis Case Finding: A Comparison of the Effectiveness of Various Roentgenographic and Photofluorographic Methods. *J Am Med Assoc* 1947;**133**(6):359.
69. Canelo C. K., Bissell D. M., Abrams H. *et al.* A multiphasic screening survey in San Jose. *Calif Med* 1949;**76**(6):409-413.

70. Clark D., Tempel C. W., Allen K. D. A. Diagnosis of pulmonary lesions discovered by mass roentgenographic survey. *J Am Med Assoc* 1950;**143**(11):943-951.
71. Lilienfeld A. M., Kordan B. A study of variability in the interpretation of chest x-rays in the detection of lung cancer. *Cancer Res* 1966;**26**(10):2145-2147.
72. Kubik A. Screening for Lung Cancer High-risk Groups. *Br Med J* 1970;**2**(5710):666-666.
73. Muhm J. R., Miller W. E., Fontana R. S. *et al.* Lung cancer detected during a screening program using four-month chest radiographs. *Radiology* 1983;**148**(3):609-615.
74. Abubakar I., Story A., Lipman M. *et al.* Diagnostic accuracy of digital chest radiography for pulmonary tuberculosis in a UK urban population. *Eur Respir J* 2010;**35**(3):689-692.
75. Balabanova Y., Coker R., Fedorin I. *et al.* Variability in interpretation of chest radiographs among Russian clinicians and implications for screening programmes: observational study. *Br Med J* 2005;**331**(7513):379-382.
76. 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):1123-1126.
77. 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.
78. 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.
79. Singh S. P., Gierada D. S., Pinsky P. *et al.* Reader variability in identifying pulmonary nodules on chest radiographs from the national lung screening trial. *J Thorac Imaging* 2012;**27**(4):249-254.
80. Doo K. W., Kang E. Y., Yong H. S. *et al.* Comparison of chest radiography, chest digital tomosynthesis and low dose MDCT to detect small ground-glass opacity nodules: an anthropomorphic chest phantom study. *Eur Radiol* 2014;**24**(12):3269-3276.
81. Litchfield D., Ball L. J., Donovan T *et al.* Learning from others: effects of viewing another person's eye movements while searching for chest nodules. Proc SPIE 2008;**6917**:691715.
82. Potchen E. J., Cooper T. G., Sierra A. E. *et al.* Measuring performance in chest radiography. *Radiology* 2000;**217**(2):456-459.
83. Robinson J. W., Brennan P. C., Mello-Thoms C. *et al.* The impact of radiology expertise upon the localization of subtle pulmonary lesions. *Proc SPIE* 2016;**9787**:97870K.
84. Robinson J. W., Brennan P. C., Mello-Thoms C. *et al.* Reporting instructions significantly impact false positive rates when reading chest radiographs. *Eur Radiol* 2016.
85. Schalekamp S., van Ginneken B., Heggelman B. *et al.* New methods for using computer-aided detection information for the detection of lung nodules on chest radiographs. *Br J Radiol* 2014;**87**(1036):20140015.
86. Daftary A. PET-MRI: Challenges and new directions. Indian J Nucl Med 2010;**25**(1):3-5.
87. Townsend D. W., Carney J. P., Yap J. T. *et al.* PET/CT today and tomorrow. *J Nucl Med* 2004;**45 Suppl 1**(1):4S-14S.
88. Kwak J. Y., Kim E. K. Ultrasound elastography for thyroid nodules: recent advances. *Ultrasonography* 2014;**33**(2):75-82.
89. Institute of Physics and Engineering in Medicine, College of Radiographers, Royal College of Radiologists. CT Equipment, Operations, Capacity and Planning in the NHS: Report from the Clinical Imaging Board. 2015 London: Clinical Imaging Board
90. Byrne S. C., Barrett B., Bhatia R. The Impact of Diagnostic Imaging Wait Times on the Prognosis of Lung Cancer. *Can Assoc Radiol J* 2015;**66**(1):53-57.
91. Royal College of Radiologists. Clinical radiology workload: guidance on on radiologists' reporting figures. 2012 London: Royal College of Radiologists
92. Royal College of Radiologists & British Society of Interventional Radiology. Provision of interventional radiology services. 2014 London: Royal College of Radiologists
93. Squires D., Anderson C. U.S. Health Care from a Global Perspective: Spending, Use of Services, Prices, and Health in 13 Countries. 2015: The Commonwealth Fund

94. NHS England. Diagnostic Imaging Dataset Statistical Release. London: Department of Health, 2016.
95. Organisation for Economic Co-operation and Development. Health at a Glance: Europe 2014. 2014: OECD
96. NHS England & NHS Improvement. Strengthening Financial Performance & Accountability in 2016/17. In: Health D. o., ed. London: NHS Improvement, 2016.
97. Cancer Strategy Taskforce. Statement of Intent: Cancer Strategy for England: 2015 – 2020. 2015 London: Department of Health
98. Cancer Research UK. CancerStats Key Facts – Lung Cancer. In: UK C. R., ed., 2011.
99. Office for National Statistics. Index of cancer survival for Clinical Commissioning Groups in England: Adults diagnosed 1998-2013 and followed up to 2014. 2016 London: Office for National Statistics
100. National Institute for Health and Clinical Excellence. Lung cancer: The diagnosis and treatment of lung cancer. 2011: National Institute for Health and Clinical Excellence 2011
101. National Institute for Health and Care Excellence. Suspected cancer: recognition and referral (NG12). 2015: National Institute for Health and Care Excellence
102. Department of Health. Radiography Skills Mix: A report on the four-tier service delivery model. 2003 UK: Department of Health
103. Price R., Miller L. An evaluation of the impact of implementation of consultant practitioners in clinical imaging. 2010 London: College of Radiographers
104. Aveyard H., Sharp P. *A Beginner's Guide to Evidence Based Practice in Health and Social Care Professions*: McGraw-Hill, 2009.
105. Dodd J. D., MacEneaney P. M., Malone D. E. Evidence-based radiology: how to quickly assess the validity and strength of publications in the diagnostic radiology literature. *Eur Radiol* 2004;**14**(5):915-922.
106. Haynes R. B. What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and consumers to pay attention to? *BMC Health Serv Res* 2002;**2**(1):3.
107. Liberati A., Vineis P. Introduction to the symposium: what evidence based medicine is and what it is not. *J Med Ethics* 2004;**30**(2):120-121.
108. Sardanelli F., Di Leo G. Introduction. In: Sardanelli F., Di Leo G., eds. *Biostatistics for Radiologists: Planning, Performing, and Writing a Radiologic Study*. Italy: Springer-Verlag Italia, 2009:1-15.
109. Sehon S. R., Stanley D. E. A philosophical analysis of the evidence-based medicine debate. *BMC Health Serv Res* 2003;**3**(1):14.
110. Wood M. A. Validation, evidence based medicine and Aristotle's wives. *J Cardiovasc Electrophysiol* 2009;**20**(7):749-750.
111. Guyatt G. H., Oxman A. D., Kunz R. *et al.* What is "quality of evidence" and why is it important to clinicians? *Br Med J* 2008;**336**(7651):995-998.
112. Guyatt G. H., Oxman A. D., Vist G. E. *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Br Med J* 2008;**336**(7650):924-926.
113. Health and Care Professions Council. Continuing professional development and your registration. 2012 London: Health and Care Professions Council
114. Ashcroft R., ter Meulen R. Ethics, philosophy, and evidence based medicine. Journal of Medical Ethics 2004;**30**(2):119-119.
115. Hek G., Moule P. *Making Sense of Research An Introduction for Health and Social Care Practitioners*. 3rd Edition ed. Great Britain: Sage, 2006.
116. Wyer P. C., Silva S. A. Where is the wisdom? I--a conceptual history of evidence-based medicine. *J Eval Clin Pract* 2009;**15**(6):891-898.
117. Rangachari P. K. Evidence-based medicine: old French wine with a new Canadian label? *J R Soc Med* 1997;**90**(5):280-284.

118. Louis P.C.A. Researches on the effects of blood-letting in some inflammatory diseases, and on the influence of tartarised antimony and vesication in pneumonitis. *Am J Med Sci* 1836;**18**:102-111.
119. Jaeschke R., Guyatt G. What is evidence-based medicine? *Semin Med Pract* 1999;**2**(3):3-7.
120. Haynes R. B., Wilczynski N.L. Optimal search stragegies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *Br Med J* 2004;**328**(7447):1040-1044.
121. Brealey S., Glenney A. A framework for radiographers planning to undertake a systematic review. *Radiography* 1999;**5**(3):131-146.
122. Devillé Walter L., Buntinx Frank, Bouter Lex M. *et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002;**2**(1):9.
123. Whiting P., Rutjes A. W., Reitsma J. B. *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;**3**:25.
124. Godfrey-Smith P. *Theory and Reality an introduction to the philisophy of science*. USA: University of Chicago Pres, 2003.
125. Guyatt G., Cairns J., Churchill D. *et al.* Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *J Am Med Assoc* 1992;**268**(17):2420-2425.
126. Blunt C. J. *Hierarchies of Evidence in Evidence-Based Medicine*. The London School of Economics and Political Science, 2015.
127. Popper K. *The Logic of Scientific Discovery*. New York: Hutchinson & Co., 1959.
128. Kuhn T. S. *The structure of scientific revolutions*. 3rd Edition ed. USA: The University of Chicago, 1996.
129. Mellor F., Knapp K. Research outcome measures. In: Ramlaul A., ed. *Medical Imaging and Radiotherapy Research Skills & Strategies*. China: Elsevier, 2010:275.
130. College of Radiographers. Code of Professional Conduct. 2013 London: College of Radiographers
131. Freedman L. S. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radiol* 1987;**60**(719):1071-1081.
132. Gatsonis C. Design of evaluations of imaging technlogy: Development of a paradigm. *Acad Radiol* 2000;**7**(9):681-683.
133. Blackmore C. C., Black W. C., Jarvik J. G. *et al.* A critical synopsis of the diagnostic and screening radiology outcomes literature. *Acad Radiol* 1999;**6 Suppl 1**(Supplement 1):S8-18.
134. Hollingworth W. Radiology cost and outcomes studies: standard practice and emerging methods. *Am J Roentgenol* 2005;**185**(4):833-839.
135. Knottnerus J. A., Dinant G., van Schaych O. P. The diagnostic before-after study to assess clinical impact. In: Knottnerus J. A., Buntinx F., eds. *The Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. 2nd ed. Singapore: Blackwell Publishing, 2009:83-95.
136. Dixon A. K. The impact of medical imaging on the physician's diagnostic and therapeutic thinking. *Eur Radiol* 1998;**8**(3):488-490.
137. Hendrick R. E., Smith R. A., Rutledge J. H., 3rd *et al.* Benefit of screening mammography in women aged 40-49: a new meta-analysis of randomized controlled trials. *J Natl Cancer Inst Monogr* 1997(22):87-92.
138. Miller A. B., Wall C., Baines C. J. *et al.* Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *Br Med J* 2014;**348**:g366.
139. Team National Lung Screening Trial Research, Aberle D. R., Adams A. M. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;**365**(5):395-409.
140. Field J., Duffy S. W., Baldwin D. *et al.* UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016;**71**(2):161-170.

141. Hardy M., Hutton J., Snaith B. Is a radiographer led immediate reporting service for emergency department referrals a cost effective initiative? *Radiography* 2013;**19**(1):23-27.
142. Ascher S. M., Shtern F. Executive summary. *Acad Radiol* 1999;**6**(Supplement 1):S3-S7.
143. Wittenberg J., Fineberg H. V., Ferrucci J. T., Jr. *et al.* Clinical efficacy of computed body tomography, II. *Am J Roentgenol* 1980;**134**(6):1111-1120.
144. Sardanelli F., Di Leo G. *Biostatistics for Radiologists: Planning, Performing, and Writing a Radiologic Study*. Italy: Springer-Verlag Italia, 2009.
145. Loop J. W., Lusted L. B. American College of Radiology Diagnostic Efficacy Studies. *Am J Roentgenol* 1978;**131**(1):173-179.
146. Lusted L. B. Decision-making studies in patient management. *N Engl J Med* 1971;**284**(8):416-424.
147. Gorry G. A., Pauker S. G., Schwartz W. B. The diagnostic importance of the normal finding. *N Engl J Med* 1978;**298**(9):486-489.
148. Ledley R. S., Lusted L. B. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* 1959;**130**(3366):9-21.
149. Kelsey Fry I. Who needs high technology? Presidential address 1983. *Br J Radiol*1984;**57**(681):765-772.
150. Thornbury J. R. Eugene W. Caldwell Lecture. Clinical efficacy of diagnostic imaging: love it or leave it. *Am J Roentgenol* 1994;**162**(1):1-8.
151. Gatsonis C. A., Hillman B. J. Introduction. *Acad Radiol* 1999;**6**(Supplement 1):S1-S2.
152. Kassirer J. P. Our suborn quest for diagnostic certainty. *N Engl J Med* 1989;**320**(22):1489-1491.
153. Schunemann H. J., Oxman A. D., Brozek J. *et al.* Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Br Med J* 2008;**336**(7653):1106-1110.
154. Springett V. H. Tuberculosis control in Britain 1945-1970-1995. *Tubercle* 1971;**52**(2):136-147.
155. Wylie C. M. Use of death rates in evaluating multiple screening. *Public Health Rep* 1961;**76**(12):1111-1116.
156. Langlotz C. P. Overcoming barriers to outcomes research on imaging: lessons from an abstract decision model. *Acad Radiol* 1999;**6 Suppl 1**(Supplement 1):S29-34.
157. Wittenberg J., Fineberg H. V. Evaluating efficacy. *Am J Roentgenol* 1980;**134**(6):1277-1279.
158. Dixon A. K., Southern J. P., Teale A. *et al.* Magnetic resonance imaging of the head and spine: effective for the clinician or the patient? *Br Med J* 1991;**302**(6768):79-82.
159. Thornbury J. R. Intermediate outcomes: diagnostic and therapeutic impact. *Acad Radiol* 1999;**6 Suppl 1**(Supplement 1):S58-65; discussion S66-58.
160. Wagner R. F. Physical performance of diagnostic imaging modalities. *Acad Radiol* 1999;**6 Suppl 1**(Supplement 1):S69-71.
161. Flamm C. R. Evaluation of appropriateness: An evidence-based technology assessment perspective. *Acad Radiol* 1999;**6**(Supplement 1):S52-S53.
162. Blanchard T. K., Mackenzie R., Bearcroft P. W. P. *et al.* Magnetic resonance imaging of the shoulder: assessment of effectiveness. *Clin Radiol* 1997;**52**(5):363-368.
163. Hobby J. L., Dixon A. K., Bearcroft P. W. P. *et al.* MR Imaging of the Wrist: Effect on Clinical Diagnosis and Patient Care. *Radiology* 2001;**220**(3):589-593.
164. Mackenzie R., Dixon A. K., Keene G. S. *et al.* Magnetic resonance imaging of the knee: assessment of effectiveness. *Clin Radiol* 1996;**51**(4):245-250.
165. Boland M. V., Lehmann H. P. A new method for determining physician decision thresholds using empiric, uncertain recommendations. *BMC Med Inform Decis Mak* 2010;**10**:20.
166. Royal College of Radiologists. A Guide to Justification for Clinical Radiologists. London: Royal College of Radiologists, 2000.
167. Department of Health. The Ionising Radiation (Medical Exposure) Regulations 2000. In: Crown, ed. London, 2000:13.

168. British Institute of Radiology, Society & College of Radiographers, Royal College of Radiologists. A guide to understanding the implications of the Ionising Radiation (Medical Exposure) Regulations in diagnostic and interventional radiology. 2015 London: Royal College of Radiologists
169. Thornbury J. R. Intermediate outcomes: diagnostic and therapeutic impact. *Acad Radiol* 1999;**6 Suppl 1**:S58-65.
170. Revicki D. A., Yabroff K. R., Shikiar R. Overcoming research in radiologic imaging: Identification of barriers and potential solutions. *Acad Radiol* 1999;**6**(Supplement 1):S29-S34.
171. Wittenberg J., Fineberg H. V., Black E. B. *et al.* Clinical efficacy of computed body tomography. *Am J Roentgenol* 1978;**131**(1):5-14.
172. Ng C. S., Watson C. J. E., Palmer C. R. *et al.* Evaluation of early abdominopelvic computed tomography in patients with acute abdominal pain of unknown cause: prospective randomised study. *Br Med J* 2002;**325**(7377):1387-1387.
173. Bedson J., Jordan K., Croft P. How do GPs use x rays to manage chronic knee pain in the elderly? A case study. *Ann Rheum Dis* 2003;**62**(5):450-454.
174. Enterline P. E., Kordan B. A controlled evaluation of mass surveys for tuberculosis and heart disease. *Public Health Rep* 1958;**73**(10):867-875.
175. Rosenkrantz A. B., Pinnamaneni N., Babb J. S. *et al.* Most Common Publication Types in Radiology Journals:: What is the Level of Evidence? *Acad Radiol* 2016;**23**(5):628-633.
176. Dixon A. K. Assessing the effectiveness of new technology. Imaging & Oncology 2007:12-17.
177. Prasad S. C., Pilepich M. V., Perez C. A. Contribution of CT to quantitative radiation therapy planning. *Am J Roentgenol* 1981;**136**(1):123-128.
178. National Institute for Health Research. National Institute for Health Research Health Technology Assessment Programme 2012 [cited 2012 12/02/2012]. Available from: http://www.hta.ac.uk/about/index.shtml.
179. Carr D., Gale A., Mugglestone M. Visual search strategies of radiographers: evidence for role extension. European Congress of Radiology. Vienna, Austria: *Eur J Radiol*, 1997:S179.
180. Manning D., Barker-Mill S. C., Donovan T. *et al.* Time-dependent observer errors in pulmonary nodule detection. *Br J Radiol* 2006;**79**(940):342-346.
181. Manning D. J., Ethell S. C., Donovan T. Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *Br J Radiol* 2004;**77**(915):231-235.
182. Habbema J. D. F., Eijkemans R., Krijnen P. *et al.* Analysis of data on the accuracy of diagnostic tests. In: Knottnerus J. A., Buntinx F., eds. T*he Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. Singapore: Blackwell Publishing, 2009:118-145.
183. Irwig L., Bossuyt P., Glasziou P. *et al.* Designing studies to ensure that estimates of test accuracy will travel. In: Knottnerus J. A., Buntinx F., eds. *The Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. Singapore: Blackwell Publishing, 2009:96-117.
184. Bree R. L. Intermediate outcomes: Diagnostic and therapeutic impact. *Acad Radiol* 1999;**6**(Supplement 1):S66-S68.
185. Brealey S., King D. G., Crowe M. T. *et al.* Accident and emergency and general practitioner plain radiograph reporting by radiographers and radiologists: a quasi-randomized controlled trial. *Br J Radiol* 2003;**76**(901):57-61.
186. Hardy M., Snaith B., Scally A. The impact of immediate reporting on interpretive discrepancies and patient referral pathways within the emergency department: a randomised controlled trial. *Br J Radiol* 2013;**86**(1021):20120112.
187. Robinson P., Wilson D. Accuracy and observer variation. *Clin Radiol* 1997;**52**(8):636-638.
188. Robinson P., Wilson D., Coral A. *et al.* Variation between experienced observers in the interpretation of accident and emergency radiographs. *Br J Radiol* 1999;**72**(856):323-330.
189. Robinson P. J. Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *Br J Radiol* 1997;**70**(839):1085-1098.

190. NHS England. Diagnostic Imaging Dataset Annual Statistical Release 2013/14. In: England N., ed. London: NHS England, 2014.
191. Anderson R. J. In search of TB cases. *Am Lung Assoc Bull* 1982;**68**(2):11-13.
192. Carroll B. E., Kurlander A. B., Nester H. G. Multiple screening pilot study: report of the Indianapolis, Indiana, project. *Public Health Rep* 1954;**69**(12):1180-1184.
193. Citron K. M. Early Diagnosis of Pulmonary Tuberculosis. *Br Med J* 1966;**1**(5487):589-591.
194. Freer A. Occurrence of Pulmonary Tuberculosis in Supposedly Screened Selectees. *Chest* 1944;**10**(3):197-209.
195. Hanaghan J. A., MacLean L. Hospital admission x-rays. 1955. *Conn Med* 2005;**69**(3):165-169.
196. Haygood T. M. Chest screening and tuberculosis in the United States. *Radiographics* 1994;**14**(5):1151-1166.
197. Holman C. B. Roentgenology in the diagnosis of tuberculosis. *Proc Staff Meet Mayo Clin* 1950;**25**(15):418-422.
198. Kurlander A. B., Carroll B. E. Case finding through multiple screening. *Public Health Reports* 1953;**68**(11):1035-1042.
199. Landau G. M., Schorsch H. A. The adult silent chest. *Radiology* 1948;**50**(1):37-43.
200. Liestol K., Tretli S., Tverdal A. *et al.* Tuberculin status, socioeconomic differences and differences in all-cause mortality: experience from Norwegian cohorts born 1910-49. *Int J Epidemiol* 2009;**38**(2):427-434.
201. Ochsner S. F. Pulmonary tuberculosis: contributions of radiology in diagnosis and treatment. *South Med J* 1986;**79**(11):1416-1424.
202. Whalen R. P., Lyons J. J. A. Medical problems of 500 prisioners on admission to a county jail. *Public Health Rep* 1962;**77**(6):497-502.
203. Hochbaum G. M. Why people seek diagnostic x-rays. *Public Health Rep* 1956;**71**(4):377-380.
204. Koppaka R., Bock N. How reliable is chest radiography? In: World Health Organisation ed. *Toman's tuberculosis: Case detection, treatment and monitoring.* 2nd Edition. Geneva: World Health Organisation, 2004:51-60.
205. Eisenberg R. L., Pollock N. R. Low yield of chest radiography in a large tuberculosis screening program. *Radiology* 2010;**256**(3):998-1004.
206. Brealey S., Scally A. J. Bias in plain film reading performance studies. *Br J Radiol* 2001;**74**(880):307-316.
207. Tudor G. R., Finlay D. B. Is there an improvement in performance when radiographs are re-reported at 24 hours? *Br J Radiol* 1999;**72**(857):465-468.
208. Brodeur A. E. Curriculum for training x-ray film readers for mass surveys. *Public Health Rep* 1954;**69**(11):1061-1063.
209. Taylor-Phillips S., Wallis M. G., Jenkinson D. *et al.* Effect of Using the Same vs Different Order for Second Readings of Screening Mammograms on Rates of Breast Cancer Detection: A Randomized Clinical Trial. *J Am Med Assoc* 2016;**315**(18):1956-1965.
210. Toman K. What were the main landmarks in the development of tuberculosis treatment? In: World Health Organisation ed. *Toman's tuberculosis: Case detection, treatment and monitoring*. 2nd Edition ed. Geneva: World Health Organisation, 2004:99-101.
211. Heelan R. T., Flehinger B. J., Melamed M. R. *et al.* Non-small-cell lung cancer: results of the New York screening program. *Radiology* 1984;**151**(2):289-293.
212. van den Bergh K. A., Essink-Bot M. L., Borsboom G. J. *et al.* Long-term effects of lung cancer computed tomography screening on health-related quality of life: the NELSON trial. *Eur Respir J* 2011;**38**(1):154-161.
213. Infante M., Lutman F. R., Cavuto S. *et al.* Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008;**59**(3):355-363.
214. 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.
215. Swinburne K. Pattern recognition for radiographers. *Lancet* 1971;**297**(7699):589-590.

216. Berman L., de Lacey G., Twomey E. *et al.* Reducing errors in the accident department: a simple method using radiographers. *Br Med J* 1985;**290**(6466):421-422.
217. McConnell J. R., Webster A. J. Improving radiographer highlighting of trauma films in the accident and emergency department with a short course of study--an evaluation. *Br J Radiol* 2000;**73**(870):608-612.
218. McConnell J., Devaney C., Gordon M. *et al.* The impact of a pilot education programme on Queensland radiographer abnormality description of adult appendicular musculo-skeletal trauma. *Radiography* 2012;**18**(3):184-190.
219. Mackay S. The impact of a short course of study on the performance of radiographers when highlighting fractures on trauma radiographs: ''The Red Dot System''. *Br J Radiol* 2006;**79**:468-472.
220. Hazell L., Motto J., Chipeya L. The Influence of Image Interpretation Training on the Accuracy of Abnormality Detection and Written Comments on Musculoskeletal Radiographs by South African Radiographers. *J Med Imaging Radiat Sci* 2015;**46**(3):302-308.
221. Smith T. N., Traise P., Cook A. The influence of a continuing education program on the image interpretation accuracy of rural radiographers. *Rural Remote Health* 2009;**9**(2):1145.
222. Piper K., Paterson A. Initial image interpretation of appendicular skeletal radiographs: A comparison between nurses and radiographers. *Radiography* 2009;**15**(1):40-48.
223. Coleman L., Piper K. Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers. *Radiography* 2009;**15**(3):196-202.
224. Melamed A., Fidler A. Routine Fluororoentgen Chest Examinations of Hospital Admissions: From the Viewpoint of the Radiologist. *Chest* 1949;**15**(3):346-353.
225. Sherman R. S. Screening of Chest Roentgenograms by Advanced Roentgen Technologists. *Invest Radiol* 1971;**6**(1):76.
226. 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.*Br J Radiol* 2001;**74**(879):230-233.
227. Manning D., Ethell S., Donovan T. *et al.* How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography* 2006;**12**(2):134-142.
228. Donovan T., Litchfield D. Looking for Cancer: Expertise Related Differences in Searching and Decision Making. *Applied Cognitive Psychology* 2013;**27**(1):43-49.
229. Chen Y., James J. J., Dong L. *et al.* Measuring performance in the interpretation of chest radiographs: a pilot study. *Clin Radiol* 2016.
230. Christiansen J. M., Gerke O., Karstoft J. *et al.* Poor interpretation of chest X-rays by junior doctors. *Dan Med J* 2014;**61**(7):A4875.
231. Zachrisson S., Vikgren J., Svalkvist A. *et al.* Effect of clinical experience of chest tomosynthesis on detection of pulmonary nodules. *Acta Radiol* 2009;**50**(8):884-891.
232. Beam C. A., Conant E. F., Sickles E. A. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* 2003;**95**(4):282-290.
233. Lusted L. B. A study of the efficacy of diagnostic radiologic procedures. Final report on diagnostic efficacy. 1977 Chicago: American College of Radiology
234. Brealey S. D., King D. G., Hahn S. *et al.* Radiographers and radiologists reporting plain radiograph requests from accident and emergency and general practice. *Clin Radiol* 2005;**60**(6):710-717.
235. World Health Organisation. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Bull World Health Organ 2001;**79**(4):373-374.
236. World Health Organisation. Handbook for good clinical research practice. 2002: World Health Organisation

237. Data Protection Act 1998. United Kingdom, 1998.
238. Tourassi G. Reciever operating characteristic analysis: basic concepts and practical applications. In: Samei E., Krupinski E. A., eds. T*he Handbook of Medical Image Perception and Techniques*. USA: Cambridge University Press, 2010:187-203.
239. College of Radiographers. Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance. 2013 London: College of Radiographers
240. Donald J. J., Barnard S. A. Common patterns in 558 diagnostic radiology errors. *J Med Imaging Radiat Oncol* 2012;**56**(2):173-178.
241. Chakraborty D. P. A status report on free-response analysis. *Radiat Prot Dosimetry* 2010;**139**(1-3):20-25.
242. Obuchowski N. A. Sample size tables for receiver operating characteristic studies. *Am J Roentgenol* 2000;**175**(3):603-608.
243. Chakraborty D. The FROC, AFROC and DROC variants of the ROC analysis. In: Van Metter R., Beutel J., Knundel H., eds. *Handbook of medical imaging*. USA: SPIE Press, 2000:771-796.
244. Chakraborty D. Recent Developments in FROC methodology. In: Samei E., Krupinski E. A., eds. T*he Handbook of Medical Image Perception and Techiques*. USA: Cambridge University Press, 2010:216-239.
245. Health and Social Care Information Centre. Hospital Episode Statistics for England. Inpatient statistics, 2011-12. 2012: Health and Social Care Information Centre
246. Chakraborty D. P. How to conduct a free-response study 2005 [Available from: http://www.devchakraborty.com/HowToConduct.html.
247. Chakraborty D. P. Clinical relevance of the ROC and free-response paradigms for comparing imaging system efficacies. *Radiat Prot Dosimetry* 2010;**139**(1-3):37-41.
248. Sim J., Wright C. C. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy* 2005;**85**(3):257-268.
249. Haygood T. M., Ryan J., Brennan P. C. *et al.* On the choice of acceptance radius in free-response observer performance studies. *Br J Radiol* 2013;**86**(1021):42313554.
250. de Lacey G., Morley S., Berman L. *Chest x-ray: a survival guide*. Spain: Elsiever, 2008.
251. Asplund S. Detection of pulmonary nodules in chest tomosynthesis: Comparison with chest radiography, evaluation of learning effects and investigation of radiation dose level dependency University of Gothenburg 2014.
252. 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.
253. Szucs-Farkas Z., Patak M. A., Yuksel-Hatz S. *et al.* Single-exposure dual-energy subtraction chest radiography: detection of pulmonary nodules and masses in clinical practice. *Eur Radiol* 2008;**18**(1):24-31.
254. Hardy M., Flintham K., Snaith B. *et al.* The impact of image test bank construction on radiographic interpretation outcomes: A comparison study. *Radiography* 2016;**22**(2):166-170.
255. Reed W. M., Ryan J. T., McEntee M. F. *et al.* The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology* 2011;**258**(3):938-943.
256. Chakraborty D. P. Validation and statistical power comparison of methods for analyzing free-response observer performance studies. Acad Radiol 2008;**15**(12):1554-1566.
257. JAFROC [program]. 4.2 version. Pennsylvania, USA, 2014.
258. Bridger D., Havutcu N. Incident reporting policy (incorporating the Serious Untoward Incident reporting policy). 2010 London: Homerton University Hospital NHS Foundation Trust
259. Havutcu N., Goodenough D., Mavers M. Being Open Policy. 2010 London: Homerton University Hopital NHS Foundation Trust
260. Care & Quality Commission. Regulation 20: Duty of candour. Care & Quality Commission. London: Care & Quality Commission, 2015.

261. Landis J. R., Koch G. G. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**(1):159-174.
262. Eisen L. A., Berger J. S., Hegde A. *et al.* Competency in chest radiography. A comparison of medical students, residents, and fellows. *J Gen Intern Med* 2006;**21**(5):460-465.
263. Eng J., Mysko W. K., Weller G. E. *et al.* Interpretation of Emergency Department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and film with digital display. *Am J Roentgenol* 2000;**175**(5):1233-1238.
264. Herman P. G., Gerson D. E., Hessel S. J. *et al.* Disagreements in chest roentgen interpretation. *Chest* 1975;**68**(3):278-282.
265. Danish Tuberculosis Index. Dual reading as a routine procedure in mass radiography. *Bull Wld Hlth Org* 1955;**12**(1-2):247-259.
266. Leslie A., Jones A. J., Goddard P. R. The influence of clinical information on the reporting of CT by radiologists. *Br J Radiol* 2000;**73**(874):1052-1055.
267. Mehrotra P., Bosemani V., Cox J. Do radiologists still need to report chest x rays? *Postgrad Med J* 2009;**85**(1005):339-341.
268. Nyboe J. Valuation of efficiency in interpretation of chest xray films. *Bull World Health Orga*n 1966;**35**(4):535-545.
269. Sakurada S., Hang N. T., Ishizuka N. *et al.* Inter-rater agreement in the assessment of abnormal chest x-ray findings for tuberculosis between two Asian countries. *BMC Infect Dis* 2012;**12**(1):31.
270. Tudor G. R., Finlay D. B. Error review: can this improve reporting performance? *Clin Radiol* 2001;**56**(9):751-754.
271. Weatherburn G. C., Ridout D., Strickland N. H. *et al.* A comparison of conventional film, CR hard copy and PACS soft copy images of the chest: analyses of ROC curves and inter-observer agreement. *Eur J Radiol* 2003;**47**(3):206-214.
272. Scally A. J., Brealey S. Confidence intervals and sample size calculations for studies of film-reading performance. *Clin Radiol* 2003;**58**(3):238-246.
273. Campbell M. K., Elbourne D. R., Altman D. G. *et al.* CONSORT statement: extension to cluster randomised trials. *Br Med J* 2004;**328**(7441):702-708.
274. Wears R. L. Advanced statistics: Statistical methods for analyzing cluster and cluster-randomised data. *Acad Emerg Med* 2002;**9**(4):330-341.
275. Ng C. S., Palmer C. R. Analysis of diagnostic confidence and diagnostic accuracy: a unified framework. *Br J Radiol* 2007;**80**(951):152-160.
276. Omary R. A., Kaplan P. A., Dussault R. G. *et al.* The impact of ankle radiographs on the diagnosis and management of acute ankle injuries. *Acad Radiol* 1996;**3**(9):758-765.
277. Tsushima Y., Aoki J., Endo K. Contribution of the diagnostic test to the physician's diagnostic thinking: new method to evaluate the effect. *Acad Radiol* 2003;**10**(7):751-755.
278. Hazard-Munro B. Selected nonparametric techniques. In: Hazard-Munro B., ed. *Statistical methods for healthare research*. 5th ed. USA: Lippincott Williams & Wilkins, 2005:109-135.
279. Littlefair S., Mello-Thoms C., Reed W. *et al.* Increasing Prevalence Expectation in Thoracic Radiology Leads to Overcall. *Acad Radiol* 2016;**23**(3):284-289.
280. Schwartz A., Elstein A. S. Clinical problem solving and diagnostic decision making: a selective review of the cogniive research literature. In: Knottnerus J. A., Buntinx F., eds. *The evidence base of clinical diagnosis: theory and methods of diagnostic research*. 2nd Edition ed. Singapore: Blackwell Publishing, 2009.
281. Brennan P. C., Ryan J., Evanoff M. *et al.* The impact of acoustic noise found within clinical departments on radiology performance. *Acad Radiol* 2008;**15**(4):472-476.
282. Kohli A., Robinson J. W., Ryan J. *et al.* Reader characteristics linked to detection of pulmonary nodules on radiographs: ROC vs. JAFROC analyses of performance. *Proc of SPIE* 2011;**7966**:79660K-79660K-79668.

283. Litchfield D., Ball L. J., Donovan T. *et al.* Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *J Exp Psychol Appl* 2010;**16**(3):251-262.
284. Buissink C., Thompson J. D., Voet M. *et al.* The influence of experience and training in a group of novice observers: A jackknife alternative free-response receiver operating characteristic analysis. *Radiography* 2014;**20**(4):300-305.
285. 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.
286. Ng C. S., Palmer C. R. Analysis of diagnostic confidence: Application to data from a prospective randomized controlled trial of CT for acute abdominal pain. *Acta Radiologica* 2010;**51**(4):368-374.
287. Committee on Diagnostic Error in Health Care. Improving Diagnosis in Health Care. 2015 Washington DC: National Academy of Sciences
288. Yamada Y., Jinzaki M., Hasegawa I. *et al.* Fast scanning tomosynthesis for the detection of pulmonary nodules: diagnostic performance compared with chest radiography, using multidetector-row computed tomography as the reference. *Invest Radiol* 2011;**46**(8):471-477.
289. Park S. H., Ahn S., Hong N. *et al.* Quality of reporting noninferiority/similarity in research studies of diagnostic imaging. *Radiology* 2014;**270**(1):241-247.
290. Egglin T. K., Feinstein A. R. Context bias. A problem in diagnostic radiology. *J Am Med Assoc* 1996;**276**(21):1752-1755.
291. Reed W. M., Chow S. L., Chew L. E. *et al.* Can prevalence expectations drive radiologists' behavior? *Acad Radiol* 2014;**21**(4):450-456.
292. Byrt T., Bishop J., Carlin J. B. Bias, prevalence and kappa. J Clin Epidemiol 1993;**46**(5):423-429.
293. Flight L., Julious S. A. The disagreeable behaviour of the kappa statistic. *Pharm Stat* 2015;**14**(1):74-78.
294. Austin J. H., Romney B. M., Goldsmith L. S. Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect. *Radiology* 1992;**182**(1):115-122.
295. Kashani H., Varon C. A., Paul N. S. *et al.* Diagnostic performance of a prototype dual-energy chest imaging system ROC analysis. *Acad Radiol* 2010;**17**(3):298-308.
296. Schalekamp S., Karssemeijer N., Cats A. M. *et al.* The Effect of Supplementary Bone-Suppressed Chest Radiographs on the Assessment of a Variety of Common Pulmonary Abnormalities: Results of an Observer Study. *J Thorac Imaging* 2016;**31**(2):119-125.
297. Gatt M. E., Paltiel O., Hilier N. *et al.* Chest radiographs in the emergency department: is the radiologist really necessary? *Postgrad Med J* 2003;**79**(930):214-217.
298. Manning D. J., Bunting S., Leach J. An ROC evaluation of six systems for chest radiography. *Radiography* 1999;**5**(4):201-209.
299. de Hoop B., De Boo D. W., Gietema H. A. *et al.* Computer-aided detection of lung cancer on chest radiographs: effect on observer performance. *Radiology* 2010;**257**(2):532-540.
300. Lee K. H., Goo J. M., Park C. M. *et al.* Computer-aided detection of malignant lung nodules on chest radiographs: effect on observers' performance. *Korean J Radiol* 2012;**13**(5):564-571.
301. Vikgren J., Zachrisson S., Svalkvist A. *et al.* Comparison of chest tomosynthesis and chest radiography for detection of pulmonary nodules: human observer study of clinical cases. *Radiology* 2008;**249**(3):1034-1041.
302. Kasai S., Li F., Shiraishi J. *et al.* Usefulness of computer-aided diagnosis schemes for vertebral fractures and lung nodules on chest radiographs. *Am J Roentgenol* 2008;**191**(1):260-265.
303. McEntee M., Quinn C. An investigation into how radiographer's comments influence doctor's decisions in a busy Dublin A&E department. European Congress of Radiology. Vienna, Austria, 2011.

304. Good B. C., Cooperstein L. A., DeMarino G. B. *et al.* Does knowledge of the clinical history affect the accuracy of chest radiograph interpretation? *Am J Roentgenol* 1990;**154**(4):709-712.
305. Berbaum K. S., Franken E. A., Jr., Dorfman D. D. *et al.* Influence of clinical history on perception of abnormalities in pediatric radiographs. *Acad Radiol* 1994;**1**(3):217-223.
306. Fuhrman C. R., Britton C. A., Bender T. *et al.* Observer performance studies: detection of single versus multiple abnormalities of the chest. *Am J Roentgenol* 2002;**179**(6):1551-1553.
307. Berbaum K., Franken E. A., Jr., Caldwell R. T. *et al.* Can a checklist reduce SOS errors in chest radiography? *Acad Radiol* 2006;**13**(3):296-304.
308. Berbaum K. S., Caldwell R. T., Schartz K. M. *et al.* Does computer-aided diagnosis for lung tumors change satisfaction of search in chest radiography? *Acad Radiol* 2007;**14**(9):1069-1076.
309. Berbaum K. S., Franken E. A., Jr., Dorfman D. D. *et al.* Role of faulty decision making in the satisfaction of search effect in chest radiography. *Acad Radiol* 2000;**7**(12):1098-1106.
310. Berbaum K. S., Dorfman D. D., Franken E. A. *et al.* Proper ROC analysis and joint ROC analysis of the satisfaction of search effect in chest radiology. *Acad Radiol* 2000;**7**(11):945-958.
311. Ng Chaan S., Palmer Christopher R. Assessing Diagnostic Confidence: A Comparative Review of Analytical Methods. *Acad Radiol* 2008;**15**(5):584-592.
312. Wood B. P., Deitte L. Building clinical decision-making skills. *Acad Radiol* 2013;**20**(11):1468-1471.
313. Buchan J., Seecombe I., Charlesworth A. Staffing matters; funding counts. 2016 London: The Health Foundation
314. NHS Confereration. Challenges facing the NHS in England: a guide for MPs and Peers. London: NHS Confereration, 2015:4.
315. Dunn P., McKenna H., Murray R. Deficits in the NHS 2016. 2016 London: The Kings Fund
316. National Imaging Clinical Advisory Group. Implementing 7 Day working inImaging Departments: GoodPractice Guidance. 2012: Department of Health
317. Imison C., Castle-Clark S., Watson R. Reshaping the workforce to deliver the care patients need. 2016 London: Nuffield Trust
318. Royal College of Radiologists. Quality assurance in radiology reporting: peer feedback. 2014 London: Royal College of Radiologists
319. Dorning H., Bardsley M. Focus on: Allied health professionals Can we measure quality of care? 2014 London: The Health Foundation and Nuffield Trust
320. Appleby J., Galea A., Murray R. The NHS productivity challenge Experience from the front line. 2014 London: The King's Fund
321. NHS Improvements. First steps towards quality improvement: A simple guide to impoving services. 2011 UK: Department of Health
322. National Patient Safety Agency. Never Events Framework:Update for 2010/11. 2010 London: National Patient Safety Agency
323. Morris S., Hunter R. M., Ramsay A. I. *et al.* Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis. *Br Med J* 2014;**349**:g4757.
324. National Peer Review Programme. National Peer Review Report: Trauma Services 2015. 2015 London: NHS England
325. Bertram R., Kaakinen J., Bensch F. *et al.* Eye Movements of Radiologists Reflect Expertise in CT Study Interpretation: A Potential Tool to Measure Resident Development. *Radiology* 2016:151255.
326. Lauritzen P. M., Stavem K., Andersen J. G. *et al.* Double reading of current chest CT examinations: Clinical importance of changes to radiology reports. *Eur J Radiol* 2016;**85**(1):199-204.

327. McGuire C. M., MacMahon P., Byrne D. P. *et al.* Diagnostic accuracy of magnetic resonance imaging and magnetic resonance arthrography of the hip is dependent on specialist training of the radiologist. *Skeletal Radiol* 2012;**41**(6):659-665.
328. Royal College of Radiologists. Unreported X-rays, computed tomography (CT) and magnetic resonance imaging (MRI) scans: Results of a snapshot survey of English National Health Service (NHS) trusts. London: Royal College of Radiologists, 2015.
329. ACE Programme. Lung Cancer Pathways: Interim Report Variations in performance and practice. 2016 London: Cancer Research UK, Macmillian & NHS England
330. NHS Improvements. The best of clinical pathway redesign: Practical examples delivering benefits to patients. 2011 UK: Department of Health
331. The Health Foundation. Fit for purpose? Workforce policy in the English NHS. 2016 London: The Health Foundation
332. Kelly B. S., Rainford L. A., Darcy S. P. *et al.* The Development of Expertise in Radiology: In Chest Radiograph Interpretation, "Expert" Search Pattern May Predate "Expert" Levels of Diagnostic Accuracy for Pneumothorax Identification. *Radiology* 2016:150409.
333. Graf B., Simon U., Eickmeyer F. *et al.* 1K versus 2K monitor: a clinical alternative free-response receiver operating characteristic study of observer performance using pulmonary nodules. *Am J Roentgenol* 2000;**174**(4):1067-1074.
334. Trafford V., Leshem S. Starting at the end to undertake doctoral research: predictable questions as stepping stones. *Higher Educ Rev* 2002;**34**(1):31-49.
335. Trafford V., Leshem S. How to make your thesis shine. European Conference for Education Research. Cadiz, Spain, 2012.
336. Leshem S., Trafford V. Overlooking the conceptual framework. *Innovat Educ Teach Int* 2007;**44**(1):93-105.
337. Trafford V., Leshem S. *Stepping stones to achieving your doctorate: focusing on your viva from the start*. UK: Open University Press, 2008.
338. Lusted L. B. Logical Analysis in Roentgen Diagnosis: Memorial Fund Lecture. *Radiology* 1960;**74**(2):178-193.
339. Knottnerus J. A., Muris J. W. Assment of the accuracy of diagnostic tests: the cross-sectional study. In: Knottnerus J. A., Buntinx F., eds. *The Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. 2nd ed. Singapore: Blackwell Publishing, 2009:42-62.
340. Metz C. E. Fundamental ROC Analysis. In: van Metter R., Beutel J., Knundel H., eds. *Handbook of medical imaging*. USA: SPIE Press, 2000:751-769.
341. Halligan S., Altman D. G., Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *Eur Radiol* 2015;**25**(4):932-939.
342. Chakraborty D., Zhai X. On the meaning of the weighted alternative free-response operating characteristic figure of merit. *Medical Physics* 2016;**43**(5):2548-2557.
343. Chakraborty D. P. New developments in observer performance methodology in medical imaging. *Semin Nucl Med* 2011;**41**(6):401-418.
344. Hanley J. A., McNeil B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**(1):29-36.
345. Obuchowski N. A., Beiden S. V., Berbaum K. S. *et al.* Multireader, multicase receiver operating characteristic analysis: an empirical comparison of five methods. *Acad Radiol* 2004;**11**(9):980-995.
346. Chakraborty D. P., Haygood T. M., Ryan J. *et al.* Quantifying the clinical relevance of a laboratory observer performance paradigm. *Br J Radiol* 2012;**85**(1017):1287-1302.
347. Schalekamp S., van Ginneken B., Koedam E. *et al.* Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images. *Radiology* 2014;**272**(1):252-261.

348. Thelle A., Gjerdevik M., Grydeland T. *et al.* Pneumothorax size measurements on digital chest radiographs: Intra- and inter- rater reliability. *Eur J Radiol* 2015;**84**(10):2038-2043.
349. Piper K. J., Buscall K. L. MRI reporting by radiographers: the construction of an objective structured examination. *Radiography* 2008;**14**(2):78-89.
350. Team DAMASK Trial. Influence of magnetic resonance imaging of the knee on GPs' decisions: a randomised trial. *Br J Gen Pract* 2007;**57**(541):622-629.
351. Tsushima Y. Diagnostic Certainty. In: Woznitza N., ed. Email ed, 2015.
352. Black W. C., Armstrong P. Communicating the significance of radiologic test results: the likelihood ratio. *Am J Roentgenol* 1986;**147**(6):1313-1318.
353. Radiology American College of. ACR Practice Guidelines for Communication of Diagnostic Imaging Findings. 2010 USA: American College of Radiology
354. Swensen S. J., Johnson C. D. Flying in the plane you service: patient-centered radiology. *J Am Coll Radiol* 2010;**7**(3):216-221.
355. Obuchowski N., Zepp R. C. Simple steps for improving multiple-reader studies in radiology. *Am J Roentgenol* 1996;**166**(3):517-521.
356. Schulz K. F., Altman D. G., Moher D. *et al.* CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br Med J* 2010;**340**:c332.
357. Kane R. L., Wang J., Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 2007;**60**(3):241-249.
358. Areia M., Soares M., Dinis-Ribeiro M. Quality reporting of endoscopic diagnostic studies in gastrointestinal journals: where do we stand on the use of the STARD and CONSORT statements? *Endoscopy* 2010;**42**(2):138-147.
359. Chauhan S. P., Berghella V., Sanderson M. *et al.* Randomized clinical trials behind level A recommendations in obstetric practice bulletins: compliance with CONSORT statement. *Am J Perinatol* 2009;**26**(1):69-80.
360. Ahn S., Park S. H., Lee K. H. How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research. *Radiology* 2013;**267**(2):328-338.
361. Piaggio G., Elbourne D. R., Altman D. G. *et al.* Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *J Am Med Assoc* 2006;**295**(10):1152-1160.
362. Andriole K. P., Prevedello L. M., Dufault A. *et al.* Augmenting the impact of technology adoption with financial incentive to improve radiology report signature times. *J Am Coll Radiol* 2010;**7**(3):198-204.
363. Goin J. E., Hermann G. A. The clinical efficacy of diagnostic imaging evaluation studies: Problems, paradigms, and prescriptions. *Invest Radiol* 1991;**26**(5):507-511.
364. Bossuyt P., Smidt N. Standards for reporting on diagnostic accuracy studies. In: Knottnerus J. A., Buntinx F., eds. T*he Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research.* Singapore: Blackwell Publishing, 2009:167-179.
365. Lijmer J. G., Mol B. W., Heisterkamp S. *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *J Am Med Assoc*1999;**282**(11):1061-1066.
366. Whiting P., Rutjes A. W., Reitsma J. B. *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;**140**(3):189-202.
367. Rutjes A. W., Reitsma J. B., Di Nisio M. *et al.* Evidence of bias and variation in diagnostic accuracy studies. *Can Med Assoc J* 2006;**174**(4):469-476.

Appendix 1 Summary of chest X-ray interpretation diagnostic accuracy for receiver operator characteristic (ROC) studies

Study	Number of Participants	Practitioner Characteristics	Number of Chest X-rays	normal:abnormal ratio	Nature of Pathology	Nature of Intervention/Comparison	AUC (average)	
Current Study	**21**	**10 consultant radiologists 11 reporting radiographers**	**106**	**53:53**	**Natural – range of pathologies**	**Direct comparison of consultant radiologists and reporting radiographers**	**CR mean AUC=0.903 RR mean AUC=0.909**	
Good *et al.*[304]	4	Consultant radiologists	247	79:168	Range of pathology (pneumothorax, interstitial lung disease, nodule)	With/without clinical history	with history range AUC=0.85–0.98	without history range AUC=0.82– 97
Berbaum *et al.*[305]	9	Consultant radiologists	64 (CXR and AXR)	32:32	General paediatric pathology	With/without clinical history	with history mean AUC=0.745	without history mean AUC=0.693
Manning *et al.*[298]	5	2 consultant radiologists 3 radiology registrars	300	150:150	Nodule detection	Different computed radiography systems (n=6)	AUC range (all observers) AUC=0.715–0.845	
Tudor and Finlay[207]	5	Consultant radiologists	50 (18 CXRs)	18:32 (CXR 4:14)	Range of pathology	Re-reporting at 24 hours	initial report mean AUC=0.88	re-report mean AUC=0.90
Eng *et al.*[263]	16	4 consultant radiologists 4 radiology registrars 4 emergency consultant 4 emergency registrars	120 (38 CXRs)	59:61	Representative of ED referrals and pathologies	Different speciality and experience PACS vs. film	PACS CR mean AUC=0.81 Rad SpR mean AUC=0.72 ED cons mean AUC=0.64 ED SpR mean AUC=0.56	Film CR mean AUC=0.85 Rad SpR mean AUC=0.78 ED cons mean AUC=0.70 ED SpR mean AUC=0.65
Potchen *et al.*[82]	162	111 consultant radiologists 29 radiology registrars 22 non-radiologist physicians	60	30:30	Representative of normal practice	Different speciality and experience	All CR mean AUC=0.860 best 20 CR mean AUC=0.953 radiology SpR AUC=0.746 non-radiologists mean AUC=0.657	
Berbaum *et al.*[309]	20	Consultant radiologists	58	29:29	Range of pathologies with/without Simulated nodule	Satisfaction of search With/without simulated nodule	no simulated nodule mean AUC=0.74	simulated nodule mean AUC=0.75
Fuhrman *et al.*[306]	8	Consultant radiologists	117	63:54	Rib fracture	Subtle fractures	multiple possible abnormalities mean AUC=0.73	single abnormality task mean AUC=0.80
Balabanova *et al.*[75]	101	13 Consultant radiologists 61 TB specialists 15 chest physicians	50	13:37	Range of pathology (Ca, infection, sarcoid) but TB weighted (n=20)	TB diagnosis	CR mean AUC=0.81 TB specialist mean AUC=0.88 chest physicians AUC=0.81	
Berbaum *et al.*[307]	20	Consultant radiologists	57	27:30	Range of pathologies with/without Simulated nodule	Satisfaction of search With/without checklist	without simulated mean AUC=0.67	with simulated mean AUC=0.68
Berbaum *et al.*[308]	16	Consultant radiologists	57	27:30	Range of pathologies with/without Simulated nodule	Satisfaction of search With/without CAD	no CAD no simulated nodules AUC = 0.681	CAD simulated nodules mean AUC=0.653

Study	Number of Participants	Practitioner Characteristics	Number of Chest X-rays	normal:abnormal ratio	Nature of Pathology	Nature of Intervention/Comparison	AUC (average)	
Kashani *et al.*[295]	5	Chest consultant radiologists	129	0:129	Nodule detection	Dual energy vs. Digital Radiography	digital radiography mean AUC=0.696	dual energy mean AUC=0.795
Kelly *et al.*[332]	20	5 consultant radiologists 4 radiology registrars 5 SHO 6 interns	30	16:14	Pneumothorax	Experience of readers	consultant radiologist mean AUC=0.947 radiology registrar mean AUC=0.792 SHO mean AUC=0.693 intern mean AUC=0.659	
Shang *et al.*[252]	4	Consultant radiologists	100	50:50	Nodule detection	Different methods of ROC calculation	mean standard AUC=0.66	

PACS = picture archiving and communication system; CXR = chest X-ray; AXR = abdominal X-ray; AUC = area under the curve; ED = emergency department; Ca = cancer; TB = tuberculosis; CR = consultant radiologist; SHO = senior house officer; CAD = computer assisted diagnosis

Appendix 2 Summary of Summary of studies which have used alternate free response receiver operator characteristic (AFROC) or jack-knife alternate free response receiver operator characteristic (JAFROC) methodology for assessment of chest X-ray diagnostic accuracy

Study	Number of Participants	Practitioner Characteristics	Number of Chest X-rays	normal:abnormal ratio	Simulated or Natural Nodules	Nature of Intervention/Comparison	Observer Performance Area Under Curve (AFROC) and Figure of Merit (JAFROC) Control	Intervention
Current Study	**21**	**10 consultant radiologists (CR) 11 reporting radiographers (RR)**	**106**	**53:53**	**Natural – range of pathologies**	**Direct comparison of consultant radiologists and reporting radiographers**	**CR mean FoM=0.786 RR mean FoM= 0.830**	
Graf *et al.*[333]	6	6 consultant radiologists	48	12:36	Natural	Low resolution vs high resolution monitors	AFROC Low resolution AUC=0.55	AFROC High resolution AUC=0.60
Manning *et al.*[227]	21	8 consultant radiologists (CR) 5 reporting radiographers (RR) before/after 6 months training 8 undergraduate (UG) radiographers (naïve)	120	40:80	?natural	Eye tracking	AFROC (naïve) UG AUC=0.63 RR before AUC=0.70	AFROC (expert) RR after AUC=0.82; CR AUC=0.80;
Kasai *et al.*[302]	18	6 consultant chest radiologists 12 consultant radiologists	60	?	Natural	With/without CAD	Without CAD Nodules FoM=0.622 Vertebral fractures FoM=0.585	With CAD Nodules FoM=0.65 Vertebral fractures FoM=0.68
Szucs-Farkas *et al.*[253]	5	4 consultant radiologists 1 radiology registrar	102	25:77	Natural	With/without digital subtraction Not just cancer – range of pathologies for nodules (sarcoid, Wegners etc.)	Without Subtraction AFROC CR1 AUC=0.559 CR2 AUC=0.595 CR3 AUC=0.446 CR4 AUC=0.506 SpR1 AUC=0.494	With Subtraction AFROC CR1 AUC=0.537 CR2 AUC=0.623 CR3 AUC=0.470 CR4 AUC=0.565 SpR1 AUC=0. 554
Vikgren *et al.*[301]	4	Consultant chest radiologists	89	47:42	Natural	Comparison with tomosynthesis	Without tomosynthesis Mean FoM=0.40	With tomosynthesis Mean FoM=0.64
Litchfield *et al.*[81]	48	24 undergraduate radiographers (UG) 24 postgraduate radiographers (PG)	42	1:1	?Simulated	Eye tracking with/without search pattern of novice/expert	Without Eye-tracking UG FoM=0.56 PG FoM=0.58	With Eye-tracking UG FoM=0.60 PG FoM=0.59
Brennen *et al.*[281]	26	11 consultant chest radiologists 15 consultant radiologists	30	1:1	Simulated	With and without noise distraction	No distraction All CR mean FoM=0.67 chest CR mean FoM=0.65 CR mean FoM=0.68	With noise distraction All CR mean FoM=0.67 chest CR mean FoM=0.68 CR mean FoM=0.69

Study	Number of Participants	Practitioner Characteristics	Number of Chest X-rays	normal:abnormal ratio	Simulated or Natural Nodules	Nature of Intervention/Comparison	Observer Performance Area Under Curve (AFROC) and Figure of Merit (JAFROC) Control	Intervention
de Hoop *et al.*[299]	6	2 consultant radiologists 4 radiology registrars	111	65:46	Natural	With and without CAD	Without CAD CR FoM=0.72 SpR FoM=0.58	With CAD CR FoM=0.93 SpR FoM=0.76
A. Kohli *et al.* [282]	12	6 consultant radiologists (CR) 6 consultant chest radiologists (CCR)	40	1:1	Simulated	ROC vs. JAFROC	all CR mean AUC=0.684 chest CR mean AUC=708 CR mean FoM= 0.659	JAFROC all CR mean FoM=0.456 chest CR mean FoM=0.487 CR mean FoM= 0.425
McEntee & Quinn[303]	4	Junior emergency physicians	42		Not given	With and without clinical history	Without History FoM=0.48	With History FoM=0.55
Yamada *et al.*[288]	3	3 consultant radiologists	116	59:57	Natural	Comparison with tomosynthesis ROC and JAFROC	Without tomosynthesis mean FoM=0.64 mean AUC=0.77	With tomosynthesis mean FoM=0.88 mean AUC=0.93
Donovan and Litchfield[228]	40	Naïve (non-medical) Undergraduate radiographers (UG) Experts (CR and RR)	30	1:1	24 natural, 4 simulated	Eye tracking study, comparison with observer experience	naïve mean FoM=0.41 1st UG mean FoM=0.60 3rd UG mean FoM=0.71 Experts mean FoM=0.72	
Doo *et al.*[80]	3	Consultant radiologists	40	10:30	Simulated	CXR vs. tomosynthesis vs. Low dose CT	CXR mean FoM=0.41	Tomosynthesis mean FoM=0.37 Low dose CT mean FoM=0.76
Littlefair *et al.* [279]	33	Consultant radiologists	47	37:10	Natural	Different anticipated prevalence	No History median FoM=0.60/0.65	Cancer history median FoM=0.64/0.70 Visa History median FoM=0.76/0.57
Robinson *et al.* [84]	10	Consultant radiologists	40	21:19	Natural	Unframed image interpretation vs. focused nodule detection	Unframed median FoM=0.671	Framed median FoM=0.571

CR = consultant radiologist; RR = reporting radiographer; UG = undergraduate radiographer; CAD = computer assisted diagnosis; AUC = area under the curve; FoM = figure of merit; CT = computed tomography; ROC = receiver operator characteristic curve

Appendix 3 Audit report of the most frequent chest X-ray diagnoses on adult inpatients at the Trust for 2011-12

Clinical Audit Report

Most frequent chest X-ray diagnoses on adult inpatients

August 2012

xxxx

Auditor(s):	xxxx
Supervisor:	xxxx
Directorate:	xxxx
Date Report Written:	August 2012

Table of Contents

1.0 Abstract/Summary

Chest X-rays (CXR) are one of the most frequently performed radiology investigations and are used in a wide range of clinical situations for diagnosis, prognosis, treatment response and surveillance. Based on activity data from the Trust, the discharge diagnosis for all adult patients who had a chest X-ray were obtained. The list was consolidated by a consultant respiratory physician. The eight most frequent diagnoses accounted for 28% of adult discharge diagnoses.

2.0 Introduction

Chest X-rays (CXR) are one of the most frequently performed radiology investigations and are used in a wide range of clinical situations for diagnosis, prognosis, treatment response and surveillance.

3.0 Aim

To identify the most common diagnoses at patient discharge in which a chest X-ray was performed during their in-patient stay for adult patients.

4.0 Method

An activity report for the financial year 2011-12 identified all in-patients who had a chest X-ray performed. The primary diagnosis for each patient was identified and summarised. A consultant respiratory physician reviewed the totals and categorised similar diagnoses into a single headline diagnosis (e.g. bronchopneumonia and lobar pneumonia combined to 'pneumonia'). The total results for each diagnosis were determined and the list of the eight most frequent compiled.

5.0 Results

Approximately 4,200 chest X-rays were performed on 2,400 adult inpatients in the financial year 2011-12. Consolidation of the discharge diagnoses and elimination of diagnoses not related to respiratory or cardiac causes, produced eight diagnoses that accounted for 695 patients, and the results are summarised in Table 1.

IP Diagnosis at Discharge	Number
Pneumonia	304
Heat Failure/ Pulm Oedema	232
Pleural Effusion	33
COPD	28
Malignancy (Primary/Secondary)	25
Perforation	29
Pneumothorax	24
TB	18
Fibrosis (Sarcoid/Rheum Arthritis)	2
TOTAL	**695**

Table 1. Most Common Diagnoses

According to the Trust annual report, there were just over 30,000 in patient occurrences during the audit period, which meant that 12.5% of patients had a chest X-ray during their stay.

6.0 Discussion

This audit demonstrates the wide range of patient presentations and clinical conditions that are investigated using a chest X-ray for adult patients. The eight most frequent diagnoses accounted for 28% of all in-patient discharge diagnoses where a chest X-ray was performed.

7.0 Conclusions

Chest X-ray are a frequently performed radiology investigation. Approximately 12.5% of hospital patients received a chest xray during their inpatient stay. The eight most frequent diagnoses accounted for 28% of all adult discharge diagnoses.

8.0 Recommendations

Chest X-ray are quick, readily accessible radiology investigation with a low radiation burden. Appropriate requesting by clinicians will help ensure that patient management and radiation dose are both optimised. Further evaluation in the next audit to include appropriateness of referrals is suggested.

9.0 Strategy for implementation

No specific recommendations.

10.0 Date for re-audit

This Clinical Audit will be re-audited on: *September 2015*.

Appendix 4 National Research Ethics Committee approval letter

Health Research Authority

21 May 2013

Dear

Study title:	**Establishing the diagnostic accuracy of radiographer chest x-ray reports and their influence on clinicians' clinical reasoning: a comparison with consultant radiologists.**
REC reference:	
IRAS project ID:	

The Research Ethics Committee reviewed the above application at the meeting held on 08 May 2013. Thank you for attending to discuss the application.

We plan to publish your research summary wording for the above study on the NRES website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the Co-ordinator

Ethical opinion

1. The Committee asked you to clarify if informed consent is required to use their x-ray images.

You explained that patient consent is not required to use x-rays and consent was obtained at the point of the x-ray being taken and is used for clinical purposes. You added that the

A Research Ethics Committee established by the Health Research Authority *patient identifiable data is removed and therefore cannot be viewed on screen.*

2. The Committee asked you to confirm that all reporting of the data will be conducted in the UK only.

You confirmed this.

3. The Committee asked you to explain why there are different reimbursement rates given for different participants in the study. Some are paid £350 and some are paid £150.

You replied that this is calculated pro rota.

4. The Committee asked you to explain what will happen if you identify an adverse finding on the x-ray and would like to inform the participant of this.

You replied that you would be able to link the x-ray back to the patient using the NHS PACS computer system.

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Ethical review of research sites

NHS Sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Decision: Favourable Opinion (with conditions)

The Committee gave a favourable opinion of the application (with additional conditions)

1)Changes to the Participant Information Sheet (PIS):

a) Please state clearly that the participant will be informed of any adverse findings found on their x-rays and have the option not to be informed of this.

2) Changes to the Consent form;

a) Please state clearly that participants have the option to be informed of any adverse findings found and provide them the option to consent to this or not.

The Committee delegated authority to confirm its final opinion on the application to the Assistant Coordinator.

You must notify the REC in writing once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which can be made available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Advertisement	1, Advertisement for Participating Reporting Radiographers	02 April 2013
Evidence of insurance or indemnity		08 April 2013
Investigator CV		01 October 2012
Letter from Statistician		26 March 2013

Other: Flowchart - Phase 1	1.0	02 April 2013
Other: Flowchart - Phase 2	1.0	02 April 2013
Other: Flowchart - Overview of Research Protocol	1.0	02 April 2013
Other: Participant Clinician Recruitment email	1.0	02 April 2013
Other: Letter from funder - CoRIPS		17 December 2012
Other: 2 x Peer Reviews		15 March 2013
Other: Patient Review - email		31 March 2013
Other: Patient Review - email		06 March 2013
Other: CV for Academic Supervisor -		
Other: CV for Academic Supervisor -		
Participant Consent Form: Consent form for Phase 1	1.0	02 April 2013
Participant Consent Form: Consent form for Phase 2	1.0	02 April 2013
Participant Information Sheet: PIS Arbiter	1.0	02 April 2013
Participant Information Sheet: PIS Clincians	1.0	02 April 2013
Participant Information Sheet: PIS Reporting Practitioner	1.0	02 April 2013
Participant Information Sheet: PIS Reference Standard	1.0	02 April 2013
Protocol	2.0	08 April 2013
REC application	3.5	11 April 2013

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

There were no declarations of interest.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

Please quote this number on all correspondence

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely

Email:

Enclosures: *List of names and professions of members who were present at the meeting and those who submitted written comments*
"After ethical review – guidance for researchers"

Copy to:

NHS
Health Research Authority

29 May 2013

Dear

Study title: **Establishing the diagnostic accuracy of radiographer chest x-ray reports and their influence on clinicians' clinical reasoning: a comparison with consultant radiologists.**
REC reference:
IRAS project ID:

Thank you for your letter of 27th May 2013. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our letter dated 21 May 2013

Documents received

The documents received were as follows:

Document	*Version*	*Date*
Covering Letter		21 May 2013
Participant Consent Form: Consent form for Phase 1	2	27 May 2013
Participant Consent Form: Consent form for Phase 2	2	27 May 2013
Participant Information Sheet: PIS Arbiter	2	27 May 2013
Participant Information Sheet: PIS Clincians	2	27 May 2013
Participant Information Sheet: PIS Reference Standard	2	27 May 2013
Participant Information Sheet: PIS Reporting Practitioner	2	27 May 2013

A Research Ethics Committee established by the Health Research Authority

Approved documents

The final list of approved documentation for the study is therefore as follows:

Document	*Version*	*Date*
Advertisement	1, Advertisement for Participating Reporting Radiographers	02 April 2013
Covering Letter		21 May 2013
Evidence of insurance or indemnity		08 April 2013
Investigator CV		01 October 2012
Letter from Statistician		26 March 2013
Other: Flowchart - Phase 1	1.0	02 April 2013
Other: Flowchart - Phase 2	1.0	02 April 2013
Other: Flowchart - Overview of Research Protocol	1.0	02 April 2013
Other: Participant Clinician Recruitment email	1.0	02 April 2013
Other: Letter from funder - CoRIPS		17 December 2012
Other: 2 x Peer Reviews		15 March 2013
Other: Patient Review - email		31 March 2013
Other: Patient Review - email		06 March 2013
Other: CV for Academic Supervisor -		
Other: CV for Academic Supervisor -		
Participant Consent Form: Consent form for Phase 1	1.0	02 April 2013
Participant Consent Form: Consent form for Phase 2	1.0	02 April 2013
Participant Consent Form: Consent form for Phase 1	2	27 May 2013
Participant Consent Form: Consent form for Phase 2	2	27 May 2013
Participant Information Sheet: PIS Arbiter	1.0	02 April 2013
Participant Information Sheet: PIS Clincians	1.0	02 April 2013
Participant Information Sheet: PIS Reporting Practitioner	1.0	02 April 2013
Participant Information Sheet: PIS Reference Standard	1.0	02 April 2013
Participant Information Sheet: PIS Arbiter	2	27 May 2013
Participant Information Sheet: PIS Clincians	2	27 May 2013
Participant Information Sheet: PIS Reference Standard	2	27 May 2013
Participant Information Sheet: PIS Reporting Practitioner	2	27 May 2013
Protocol	2.0	08 April 2013
REC application	3.5	11 April 2013

You should ensure that the sponsor has a copy of the final documentation for the study. It is the sponsor's responsibility to ensure that the documentation is made available to R&D offices at all participating sites.

	Please quote this number on all correspondence

Yours sincerely

E-mail:

Copy to:

Appendix 5 Research and Development approval from the Trust

19 June 2013

Dear

Re: Research Study: "Accuracy and diagnostic impact of radiographer chest x-ray reporting".

R&D No:

Thank you for sending all the relevant documents for Research and Development Approval of the above research study. As part of the Research and Development approval process we have conducted a site specific assessment for this study. I am happy to inform you that the Trust has approved the conduct of the study and that the Trust will indemnify against negligent harm that might occur during the course of this project.

The following main documents have been received by R&D department as part of the approval process;

Protocol Version 2	Dated: 08/04/2013
Patient Information Sheet Version 2	Dated: 27/05/2013
Consent Form Version 2	Dated: 27/05/2013

All other documents that you are required to submit as part of the process have been received.

I would like to draw your attention to the following conditions of the approval of this research project with which you must comply. **Failure to do so may result in the Trust withdrawing R&D approval which allows you to conduct this research project at**

Untoward events - Should any untoward event occur it is **essential** that you complete a clinical incident form and write on the form '**R&D**'. Contact the R&D Office immediately and if patients or staff are involved in an incident you must also contact the Risk Manager on

Status of Research - Inform us if your project is amended or if your project terminates early/requires an extension as well as informing the Research Ethics Committee. This is necessary to ensure that your indemnity cover is valid and also helps the office to maintain up-to-date records. A copy of any publications arising from the research should be sent to the R&D Office for use in the R&D Annual Report. Please be reminded that this hospital should be acknowledged in any publication.

First Patient Recruited Within 30 Days - The Department of Health (DoH) expects the first patient should be recruited within 30 days from the date of this letter. The Trust has to submit quarterly report to the DoH on this key performance Indicator (KPI). Failure to meet this KPI will result in the DoH withdrawing the funding Trust receives to support research at Homerton. R&D will contact you shortly to see if you have met this KPI.

Research Information - You will be required to complete a project update as required by the R&D Office to ensure that we have up to date information so that we can send accurate reports to the DoH and research networks. The project update form will be emailed or sent to you by the R&D Office.

Research Governance - As part of research governance, all investigators accessing identifiable personal information are required to comply with current data protection requirements.

Intellectual Property - If you believe that protectable intellectual property may arise from your research, please contact the who will advise you on the proper course of action.

Monitoring of Studies – You must comply with the Trust's legal responsibility as host of this research project to monitor and audit the research to ensure that the Research Governance Framework and Good Clinical Practice (GCP) if applicable is being adhered too. Monitoring questionnaires will be sent to you and random audit visits will also take place across the trust and will be conducted following at least a seven day notice period. **Failure to respond to any of these monitoring or auditing requests may result in the Trust withdrawing your R&D approval to conduct this research at**

Please note that all NHS and social care research is subject to the DoH *Research Governance Framework*. If you are unfamiliar with the standards contained in this document, you may obtain details from the Trust R&D Office or from the DoH website (www.dh.gov.uk).

Please do not hesitate to contact Research and Development Manager or me if you have any further questions.

Yours sincerely,

Appendix 6 Participant Information Sheet – Reference Standard Expert Chest Radiologist

Dear Colleague,

We are conducting a study which will examine the accuracy of chest x-ray interpretation of consultant radiologists and reporting radiographers.

The nature of the study mandates that a robust reference standard diagnosis is known for each chest x-ray that will be used within the image bank. It is a requirement that the image bank consists of chest x-rays that include normal and abnormal images with a range of pathologies. As an expert consultant radiologist with an interest in thoracic imaging you are invited to form part of the team that will establish the reference standard diagnosis for this image bank.

The study protocol requires 106 cases to be included within the image bank. The literature on agreement in chest x-ray interpretation between consultant radiologists estimates that approximately 220 cases would need to be interpreted. The vast majority of the cases will be randomly selected from chest x-rays obtained for clinical reasons from a single acute district general hospital in London.

You will have access to all the imaging available for each case, including follow up chest x-rays and cross sectional imaging (CT, US and MRI). You will be asked, independently and blinded to the clinical report, to indicate:

- If the chest x-ray is normal or abnormal
- To provide a free text report of the image, including all salient findings, a diagnosis (if appropriate) and any specific recommendations
- Allocate all abnormal cases to a broad disease category (for case stratification)
- To list and localise all abnormalities on the chest x-ray
- To assign a conspicuity rating

Image interpretation will occur over a minimum of two sessions, although the initial session will be the longest. It is estimated that 8 hours would be required for the first session, with follow up session(s) of 3 hours as required to reach the designated number of cases. Image interpretation will occur in the Radiology Department of the the Trust and you will be given familiarity training by the chief investigator on the use of PACS, voice recognition and the study proforma.

You will be assigned a unique identifier for the study and all of your answers will remain confidential. If you wish, you can be provided with the summary performance and your individual results after the study has been completed.

In acknowledgement of the time and effort required to participant in the study an honorarium of £350 will be paid, in addition to reasonable travel expenses.

If you have any questions or to express interest in participating, please contact the chief investigator, xxxx, on xxxx or xxxx .

Appendix 7 <u>Participant Information Sheet – Participant Consultant Radiologists & Reporting Radiographers</u>

Dear Colleague,

We are conducting a study which will examine the accuracy of chest x-ray interpretation of consultant radiologists and reporting radiographers.

The study requires 106 cases within the image bank which have been randomly selected from chest x-rays obtained for clinical reasons from a single acute district general hospital in London.

You will be provided with the clinical information provided at the time of initial request and access to previous chest x-rays (if available). You will be asked, independently and blinded to the clinical report:

- To provide a free text report of the image, including all salient findings, a diagnosis (if appropriate) and any specific recommendations
- To assign a confidence rating to each abnormality

Image interpretation will occur over two sessions on a single day, separated by an hour lunch break. Image interpretation will occur in the Radiology Department of the Trust. On the day of the study you will be given training by the chief investigator on the use of PACS, voice recognition, the study proforma and use of the confidence scale for abnormalities.

You will be assigned a unique identifier for the study and all of your answers will remain confidential. If you wish, you can be provided with the summary performance from the study and your individual results after the study has been completed.

In acknowledgement of the time and effort required to participant in the study an honorarium of £150 will be paid, in addition to reasonable travel expenses.

If you have any questions or to express interest in participating, please contact the chief investigator, xxxx, on xxxx or xxxx .

Appendix 8 Participant Information Sheet – Arbiters

Dear Colleague,

We are conducting a study which will examine the accuracy of chest x-ray interpretation of consultant radiologists and reporting radiographers.

The study requires 106 cases within the image bank which have been randomly selected from chest x-rays obtained for clinical reasons from a single acute district general hospital in London.

In order to produce a robust reference standard for the image bank, two expert chest radiologists will independently report each case. Cases will only be included if both experts are in agreement. The expert chest radiologists will be asked to localise, diagnose and list all abnormalities on the x-ray, assign a broad suspected disease category and conspicuity score. Based on the literature, it is estimated that this will require approximately 220 cases to be interpreted.

Ten consultant radiologists and ten reporting radiographers will interpret the chest x-ray image bank independently, blinded to the reference standard diagnosis and the clinical report. They will be asked to produce free-text reports which contain the localisation of every abnormality and to rate each lesion with a confidence score (1-4). These ratings will be included within the free-text report.

As an independent arbiter you will be required to compare the abnormality location lists of the two expert chest radiologists who will form the reference standard, approximately 220 cases. You will also be asked to compare the consultant radiologist and reporting radiographer chest x-ray reports against the reference standard diagnosis, a total of 2120 reports, and extract the confidence rating assigned to each abnormality.

Training in the acceptance criteria will be provided by the chief investigator.

Report comparison will be performed over multiple sessions, both when obtaining the reference standard and when the reporting radiographer and consultant radiologist reports are compared to the reference standard.

You will be assigned a unique identifier for the study and all of your answers will remain confidential.

Reasonable travel expenses will be paid.

If you have any questions or to express interest in participating, please contact the chief investigator, xxxx, on xxxx or xxxx .

Appendix 9 <u>Participant Information Sheet – Clinician</u>

Dear Colleague,

We are conducting a study which will examine the influence that chest x-ray reports have on clinicians' diagnostic thinking.

Participant clinicians will be given 27 concise case summaries (approximately 1/3 of a page) which will contain all of the relevant clinical history, physical examination findings and laboratory results. You will be asked to indicate your most likely and most important diagnosis for each case from a predefined list and to assign a confidence score for each diagnosis.

Following a washout period you will be presented with the same case summaries in conjunction with a chest x-ray report from the image obtained for that patient. A total of 265 case report pairs will be presented, 27 per week over a period of 10 weeks. You will be asked to complete the same diagnosis and confidence measures as previously. It will take approximately 90 minutes per week to complete the 27 case summaries.

You will be assigned a unique identifier for the study and all of your answers will remain confidential.

In acknowledgement of the time and effort required to participant in the study an honorarium of £150 will be paid.

If you have any questions or to express interest in participating, please contact the chief investigator, xxxx, on xxxx or xxxx .

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

Appendix 10 Consent Form – Reference Standard, arbiters, Reporting Practitioners

Patient Identification Number for this trial: RS*n*, A*n*, R*n*

CONSENT FORM

Title of Project: **Accuracy and diagnostic impact of radiographer chest x-ray reporting**

Name of Researcher: **xxxx**

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated **2nd April 2013** (version **1.0**) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant data collected during the study, may be looked at by individuals from the research team from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my responses. ☐

4. I agree to take part in the above study. ☐

______________________	______________________	______________________
Name of Participant	Date	Signature
______________________	______________________	______________________
Name of Person taking consent.	Date	Signature

Consent form version 1.0 2013_04_02

Appendix 11 Consent form – Clinician

Title of Project: **Accuracy and diagnostic impact of radiographer chest x-ray reporting**

Name of Researcher: **xxxx**

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated **2nd April 2013** (version **1.0**) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant data collected during the study, may be looked at by individuals from the research team from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my responses. ☐

4. I agree to take part in the above study. ☐

______________	______________	______________
Name of Participant	Date	Signature
______________	______________	______________
Name of Person taking consent.	Date	Signature

Consent form version 1.0 2013_04_02

Appendix 12 Conceptual framework and methodological justification

Several conceptual frameworks could be used to investigate radiographer chest X-ray reporting, with several methodologies for diagnostic accuracy and influence that reports have on clinicians' diagnostic decision-making. Appendix 12 outlines the conceptual framework used for the study. A critical review of diagnostic accuracy and diagnostic decision-making methodologies was performed. The strengths and weaknesses of each study design have been identified, and the methodologies for Part 1 (diagnostic accuracy) and Part 2 (diagnostic decision-making) discussed and justified. This has been included as an appendix to preserve flow between the literature review, identification of a gap in the current evidence base and the research questions and aims of the study.

12.1 Conceptual Framework

The notion of a conceptual framework is fundamental to all doctoral research.[334 335] It is this concept that provides structure for research questions, identifies the appropriate investigative pathway and enables the theoretical results to be transformed into real-world knowledge that, in turn, can be used to inform practice.[336 337]

The evidence based medicine/evidence based practice paradigm requires that all diagnostic and treatment decisions are based on the best, most robust, evidence available in conjunction with patient preferences, and incorporating clinician/practitioner experience.[59 106] The evidence used to underpin these decisions and support practice should ideally be based on studies that have established an improvement in patient outcome. As detailed in Chapter 2.2 and 2.3, while randomised controlled trials are often the most robust study design, other methodologies are often more appropriate in a diagnostic setting.

Alternative study designs to evaluate radiology investigations have been proposed, beginning in the 1960s and 1970s, and culminating in the seminal piece of Fryback and Thornbury,[57] who outlined, defined and described a six-tier hierarchy of efficacy for the evaluation and implementation of new healthcare technology. Various approaches used to assess efficacy can be placed within a conceptual framework, and the hierarchy enables different technologies to be evaluated and compared. Intermediate outcome measures have been developed to examine the sequential efficacy levels,[159 165 184] with this model forming the basis of Health Technology Assessment.[59]

12.2 Role of the hierarchy of efficacy in designing studies of reporting radiographer research

As outlined in Chapter 2.6 the hierarchy of efficacy is the conceptual framework (Appendix 12.3) used when assessing health technologies,[59] that includes radiographer reporting. The six-tiered hierarchy developed by Fryback and Thornbury[57] and adapted for radiographer reporting by Brealey and Scally[63] has informed the design of this study.

This robust model of assessing health technologies, including imaging and radiographer reporting, has been rigorously tested during the evaluation of multiple new modalities,[56 158 164 171] is the mandated standard for the National Institute for Health and Clinical Excellence[60] and is the approach advocated by the National Institute of Health Research.[178] The Health Technology Assessment Programme has made the use of this framework as an essential component of healthcare technology evaluation[60 178] and an assessment of chest X-ray reporting by radiographers is within the framework.

12.3 Conceptual framework for the current study

Traditionally, medical image interpretation and the provision of a clinical report has been the domain of the consultant radiologist as detailed in Chapter 2.6. Radiographer reporting is an alternative method of imaging interpretation, and is being used with increasing frequency to increase diagnostic capacity in response to growth in the volume and complexity of radiology investigations. As a method of diagnosis, and as an alternative to an existing technique (consultant radiologist interpretation), radiographer reporting was evaluated as a new health technology in the current study in line with the established framework.[57 60-63] Following the established health technology assessment framework,[59] the hierarchy of efficacy developed by Fryback and Thornbury,[57] revised for the United Kingdom by Mackenzie and Dixon[61] and adapted for radiographer reporting by Brealey[62] was used as the theoretical framework for the current study. The intermediate outcome measures utilised in the study relate to level 2 and level 3 efficacy in the hierarchy, namely diagnostic accuracy and influence on clinicians' diagnostic decision-making (diagnostic impact).[57 61 62]

12.4 Methodology

Two studies have been performed which investigate different but related aspects of radiographer reporting of adult chest X-rays. A positivist approach was used, and quasi-experimental diagnostic accuracy and diagnostic impact studies conducted.

12.4.1 Methodology for the diagnostic accuracy study

Measures of diagnostic performance include measures such as accuracy,[63 129] error rate[182] and yield.[145 338] While these give a broad picture of observer performance they do not communicate or differentiate between false and true negatives and positives. This is vital information when assessing the clinical usefulness of a test, with different consequences depending on the situation, for example the relative harm of a false negative diagnosis in a screening scenario.[155] Sensitivity and specificity offer further indicators of performance, the relative proportion of false negative and false positive results respectively.[131 144 339] This information is vital for clinicians when interpreting the results of a diagnostic investigation; a negative result for a test with high sensitivity is reassuring. A test, which produces a high number of false positives in order to achieve high sensitivity, resulting in unnecessary further investigation or treatment, may not be suitable for use in clinical practice.[105]

Positive and negative predictive values incorporate disease prevalence with the sensitivity and specificity of a test and enable clinicians to interpret the results for individual patients within a given population.[45 182 247] Predictive values allow clinicians to estimate the pre- and post-test likelihoods,[182] that enables the combination of multiple estimates from a range of sources, such as clinical signs, patient history, laboratory and radiological investigations.[165 182] This method of diagnostic decision-making incorporates thresholds, the level at which clinicians will begin or withhold treatment as they have achieved sufficient reassurance that disease is present/absent.[165] If these thresholds are not met, this will prompt the clinician to gain further information, in the form of investigations (including radiology).[165]

12.4.1.1 Receiver operator characteristic (ROC) curve methodology

Receiver operator characteristic (ROC) curves incorporate the use of thresholds and acknowledge uncertainty exists in diagnostic decisions. Adapted from signal detection theory,[146 338] ROC analysis recognises that the decisions in image interpretation in clinical practice are rarely fixed normal/abnormal thresholds; observers can alter this diagnostic threshold and influence the sensitivity and specificity of a test.[340 341] This relative trade-off between sensitivity and specificity in ROC analysis occurs as only true positive and false positive events contribute to observer performance.[341 342] Observers in ROC tasks are required to summarise the entire image and assign a confidence rating to all abnormal cases, that is to rate as abnormal an image in which the ‘signal’ (abnormality) is greater than the ‘noise’ (normal structures).[242 247 298 341] Several different rating scales (4 point, 5, point, 7 point) have been also been proposed.[340 343 344] Several assumptions are required for ROC analysis; the need for the ‘truth’ regarding each diagnosis to be known, that the observer must maintain the same confidence scores across the entire task and the bank contains both normal and abnormal cases.[242 243 345]

Observer performance using the ROC methodology is determined by the area under the ROC curve (AUC); observers who confidently identify normal and abnormal cases will have a larger AUC up to a maximum of 1 (all TP cases identified with maximum confidence, no FP cases).[298 340] There are several well-recognised limitations in the ROC paradigm, centred on the forced binary normal/abnormal decision for the entire image.[64 243 341 344] The ‘right for the wrong reason’ paradox, where a false positive and false negative diagnosis cancel each other has been discussed, and images with multiple abnormalities cannot be assessed.[343] Receiver operator characteristic curves also assign equal weight to sensitivity (true positive) and specificity (true negative) decisions.[344] In many situations such as screening, these decisions will not have equal clinical importance.[64 243 341] There is debate in the literature about the significance of these limitations; some argue that providing the lesion is rated abnormal further investigation or treatment will occur and the patient will come to no harm while others disagree, with these errors contributing to inflated observer performance that does not reflect clinical practice.[64 341] It is essential that research conducted in a controlled setting, and this includes diagnostic accuracy research, replicates and reflects clinical practice as closely as possible. The clinical significance of the ‘right for the wrong reason’ paradox is significant and could potentially result in patient harm, due to incorrect or unnecessary treatment. For this reason, the free response paradigm was used, and these considerations incorporated into the study design. This ensured that the results of the study would be an accurate reflection of reporting practitioner performance and the results generalisable to clinical practice.

12.4.1.2 Free response receiver operator characteristic (FROC) methodology

Free response ROC (FROC) is an evolution on ROC analysis that requires that each lesion is to be identified, localised and assigned a confidence rating.[241] An acceptance radius for a true or false positive decision needs to be defined, and is based on both clinical and technical parameters.[64 241 249] The free response paradigm is more representative of clinical practice; the number of possible abnormalities is unknown to the reader and images with multiple abnormalities can be assessed.[243 244] Another important distinction between FROC and ROC is that as there is no forced binary normal/abnormal decision, each FP event contributes to observer performance.[64 241 243] The measure of observer performance is also different in ROC and FROC analyses; FROC studies employ a figure of merit (FoM) instead of the AUC due to statistical variations (the FoM is not contained within a unit square [0 – 1], rather continues to infinity) due to the unlimited number of false positive decisions.[241]

An important assumption for ROC and FROC methodologies is that each observation (TP or FP) is an independent event. The contribution that multiple FP events on a single case have on observer performance in FROC analysis violates this assumption; if the signal to noise threshold has been reached (corresponding to the highest rated FP event), then by definition all other FP events are not truly independent.[243] This shortcoming has been addressed with alternate free response receiver operator characteristic (AFROC) curve analysis. With this method of analysing free response data, only the highest rated FP contributes to observer performance.[243] Another limitation of FROC analysis which is addressed with AFROC is the unequal weight given to cases with more than one lesion/abnormality compared to abnormal lesions with a single abnormality.[342]

Although AFROC is suitable, jack-knife alternate free response receiver operator characteristic (JAFROC) is at the leading edge of observer performance methodologies.[241 346] Building on the strengths of AFROC, JAFROC calculates pseudo-values by sequentially removing each case from analysis and recalculates performance that enables the performance for each case to be determined, resulting in increased statistical power.[247] Another important feature of AFROC analysis, in particular JAFROC, is the ability to assign relative weights to each lesion/abnormality contained on an abnormal case. Traditional ROC and FROC analysis assigns equal weight to each abnormality on an abnormal image regardless of the clinical significance and rewards the observer equally for each abnormality detected.[342] This limitation is overcome with AFROC analysis; each abnormality can be assigned relative weight based on the clinical significance of each lesion without skewing overall observer performance for cases with multiple abnormalities compared to a single abnormality.[342] In weighted JAFROC analysis, each lesion is assigned a weight, with the sum of all abnormalities included in a case equal to 1.[247 342] The current study therefore utilised the alternate free response methodology analysed using the JAFROC method to investigate the diagnostic accuracy of the radiographer and radiologist chest x-ray interpretations.

12.4.2 Reference standard diagnosis for the diagnostic accuracy study

The importance of a robust reference standard has been unanimously recognised in the literature as a fundamental requirement for effective analysis of the diagnostic accuracy of medical tests.[122 246] More recently, the fallibility of long held 'gold standards', such as histological diagnosis has been identified and also that invasive tests are not appropriate for all diagnostic accuracy studies.[45 183] Variation within the reference standard diagnosis is now recognised as a potential source of bias in diagnostic accuracy studies within the methodological literature.[45 183 339] Previously held to be the gold standard of image interpretation, the use of a single consultant radiologist opinion as the reference standard diagnosis.[233 276] is a failure to acknowledge the considerable body of evidence describing observer variation in image interpretation.[77 82 188 189 214] Other considerations must also be incorporated into the choice of reference standard. Invasive tests and additional radiation exposure raise significant ethical concerns for diagnostic research; subjecting patients to these risks without due need is unacceptable.[45] One method described to avoid the need for an invasive reference standard diagnosis is the use of expert clinical panel consensus and follow up.[246 339]

The most appropriate reference standard diagnosis will vary depending on the nature of the study; histological confirmation is traditionally used as the reference standard diagnosis for cancer,[77] although the use of another modality, most commonly computed tomography, is also widely reported[252 299 347] and microbiological results (culture) are frequently used for studies which examine the diagnosis of tuberculosis.[74 269] For studies which include a range of pathologies, images are often taken from the teaching files of experienced radiologists[262] which may or may not incorporate clinical case note review[348] or the use of another diagnostic modality.[82 230] So that a broad disease spectrum was included in the current study without the need for further investigations, the reference standard diagnosis was taken to be the consensus decision of two expert consultant radiologists with a subspecialist interest in thoracic imaging. Each expert assessed all available imaging independently, blinded to the initial clinical interpretation made at time of examination. These interpretations were assessed for concordance by two experienced arbiters with experience in assessing radiological reports for agreement in both a clinical and academic setting. The use of multiple, independent expert interpretation as a reference standard in radiology diagnostic accuracy studies, typically two or three consultant radiologists, is well established. The staged approach used by Piper *et al.* had all cases reviewed by a consultant radiologist with discordant cases sent for review by an independent second radiologist,[22] and comparable to the approach used for constructing an objective structured examination for magnetic resonance reporting.[349] Robinson *et al*. found agreement fell from approximately 80% for between two consultant radiologists to 61% when three consultant radiologists interpreted a case.[38] This variation was considered by Piper *et al*., who only included skeletal X-rays in an image bank when three consultant radiologists were in agreement.[23] In these studies, observers only had access to the index examination, whereas the expert chest radiologists in the current study also had access to all previous and follow up imaging available, including computed tomography. This additional information has ensured that a robust reference standard diagnosis was obtained. The final clinical diagnosis was made by case note review by a professor of medicine, specialised in respiratory medicine. This review, taken at 18-24 months after the initial X-ray examination, synthesised all available information and was correlated with the radiological diagnosis. The consensus diagnosis of two expert thoracic radiologists consisted of a list of all abnormalities on each X-ray, localisation and diagnosis. Precise localisation and diagnosis of all abnormalities enabled the diagnostic accuracy of participants to be assessed using the alternate free response approach. The clinical case note review facilitated a robust clinico-radiological diagnosis for each case.

12.4.3 Diagnostic Impact

Diagnostic impact, or diagnostic thinking efficacy, is an intermediate outcome measure that determines the influence that a radiology report has on the diagnostic decision-making of clinicians.[57 61 62] Several measures have been proposed, which include change in clinician diagnosis,[164 350] change in confidence of an existing diagnosis,[176] exclusion of a diagnosis from the differential diagnosis list[135 143 147] and displacement of other investigations.[56 170]

Two methodologies have been used in the literature to measure diagnostic impact. The methodology employed by Lusted was structured such that pre and post X-ray diagnostic confidence was assessed.[233] The study design required the treating clinician to specify a most likely and most serious diagnosis for each patient, and to determine the diagnostic confidence in each decision on a 0 – 100 scale. These assessments were performed both prior to and then in conjunction with the X-ray report. The pioneering work of Fineberg and colleagues that quantified the contribution that CT reports had on clinicians' diagnostic thinking differed from this approach. As an alternate to direct measures of diagnostic confidence they employed indirect measures of diagnostic confidence; confirmed a suspected diagnosis, reduced further investigation, the unexpected normal examination.[56 143] A limitation to the indirect measure of diagnostic impact utilised by Fineberg *et al*. is that this could fail to capture a change in diagnosis which should have been attributed to the CT report (reduces further investigation).[56] Subsequent work that examined the diagnostic efficacy of MRI by Dixon and colleagues again employed the more direct approach, pre and post-test diagnoses and diagnostic confidence to quantify the influence that MRI reports had on clinicians' diagnostic decision-making.[158 162-164]

The use of pre and post imaging diagnostic confidence enables analysis that is more robust, the influence that the radiology report has on decision-making can be quantified. Various permutations of this have been used; Lusted asked clinicians for the most likely and most serious diagnosis for emergency X-rays.[233] Only the most likely diagnosis was required for the analysis of emergency ankle X-rays by Omary *et al*.[276] and the analysis of MRI by Dixon and colleagues.[136 158 162-164] The decision to limit clinician choice to the two main diagnoses (most likely/most serious), or to prune the decision tree,[146 233] is acceptable when determining diagnostic impact and is compatible with medical decision-making theory.[165] Considering the issues discussed above, the methodology employed in the current study required clinicians' to specify a diagnosis and diagnostic confidence for their most likely and most serious diagnoses for each case and was the most appropriate study design. This is the established method when assessing diagnostic impact[276] as it provided a direct and quantifiable measure of the influence that consultant radiologist and reporting radiographer chest X-ray reports have on clinicians' diagnostic decision-making. This facilitated comparison between the practitioners who provide the reports.

The influence that radiology investigations have on clinician diagnostic confidence is quantified in several ways. A new post investigation diagnosis indicates a change in clinician diagnostic thinking. A change in confidence in an existing diagnosis is assessed by comparing the confidence levels pre and post examination, using either continuous or categorical measures.

Previous work that examined diagnostic impact failed to consider the accuracy of the clinicians' final diagnosis, an important limitation with the potential to alter patient outcome significantly.[56 233] Lusted used the interpretation of a single consultant radiologist as the sole measure of accuracy, as "case note review seldom proved this diagnosis to be incorrect"(p.175).[233] The same method was employed in the evaluation of MRI.[162-164] This is in stark contrast to the literature on image interpretation, as both historic and contemporary studies highlight the inherent variation in medical image interpretation. This constraint was identified in recent work that examined the influence that abdominal-pelvic CT has on diagnostic thinking, as Ng *et al*. suggested that there is nothing more dangerous than a clinician who is confidently wrong.[275]

Methods to correct for an incorrect radiology diagnosis have been developed.[275-277] While these have focused on CT examinations, the structure is transferrable to any radiology investigation. The final, definitive diagnosis for each case is obtained, which combines clinical follow up, laboratory, surgical and histology results.[275 277] When a radiology report is associated with an incorrect clinician diagnosis, using the Tsushima method, the diagnostic confidence is corrected (turned into a negative value).[277] Correction of confidence recognises the potential detrimental impact on patient management and outcome associated with the incorrect diagnosis.[277] This correction also acknowledges the threshold nature of diagnostic and therapeutic decisions; incorrect radiology reports that would move clinicians' confidence above the treatment threshold are more heavily penalised, reflecting the increased detrimental effect on patient outcome.[165] This is also true in the converse situation, when disease has been made very unlikely, as in the incorrect normal report. The accuracy of a clinicians' final diagnosis is intimately related to patient outcome and of fundamental importance when diagnostic confidence is to be used as an intermediate outcome measure. In order to incorporate this vital aspect into the study design, the Tsushima methodology has been employed in the current study when the diagnostic influence of chest X-ray reports on clinicians' diagnostic decision-making was examined. The Tsushima methodology, to date, has only been used in studies performed in clinical practice. The author has indicated that this method of assessment should be transferable to a structured assessment (personal communication).[351]

Criticism of diagnostic impact studies include the incomplete reporting of the methods used; it is difficult to critique a study where the methods have not been fully reported.[133] Guidance on the design, conduct and reporting of observational studies have been developed, the STengthening the

Reporting of OBservational studies in Epidemiology (STROBE) statement.[49] Concerns have also been raised about the external validity of diagnostic impact research, and it is suggested that the results are only applicable to the study population (patient and clinician).[133] The use of multiple cases, each reported by a range of reporting practitioners (reporting radiologists and consultant radiologists), assessed by multiple clinicians of different experience (consultants, specialist registrars, junior medical staff) enabled the current study to have improved external validity when compared to previous work.

The initial studies that examined the influence that CT had on diagnostic thinking when it was first implemented found that the impact on clinicians thinking increased over the study period. As the technology became more established and embedded in patient care, clinicians became aware of the limitations and benefits of the information provided by the radiology investigation and adapted their practice accordingly.[143] With experience of this new technology, clinicians were able to synthesise data with established investigations, clinical examination, and patient information to formulate a diagnosis that incorporated all of this information. This is not an issue with chest X-rays, as they are not a new technology rather an established diagnostic tool that has become a key component of clinical practice.

The clinician has the critical role in requesting, gathering and combining multiple sources of information to formulate a diagnosis for each patient. As such, the clinician has an important, central role in the diagnostic process. The radiology report is often the sole method of communicating the information provided by a radiology investigation to the clinician. It is imperative that information produced by radiology in the form of written radiology reports is communicated effectively to clinicians. If radiology report are to have an influence on patient management there are two outcomes required of a radiology report. Firstly, that clinicians understand and recognise the importance of findings and that clinicians are able to incorporate the information into diagnostic process.[352-354] As such, an essential component of all radiology reports is that the information is presented in a way that clinicians can understand, recognise and use in patient diagnostic and treatment decisions. In order for diagnostic decision-making to be influenced, the findings must be communicated effectively. The second part of the study assessed how the diagnostic decision-making of clinicians is influenced by the reporting radiographer and consultant radiologist chest X-ray reports.

12.5 External Validity

Image selections plays a crucial role in determining the external validity of diagnostic accuracy studies.[206] Highly selective banks, which include complex and/or rare pathologies can discriminate between observers with a high degree of confidence. This atypical setting will rarely reflect clinical practice.[63 206] Conversely, selecting a consecutive series from routine care will often result in a high proportion of normal or obviously abnormal cases that will reduce the ability of subsequent analysis to identify small yet important differences in observer performance.[63] For example, an uncommon yet important pathology may be deliberately included in a selected image bank to ensure that the practitioner under examination will recognise this thus ensuring significant findings are not missed in routine care but this may not occur in a consecutive series of images drawn from clinical practice.[206] It is a fine balance; the choice should be made by determining the focus of the study.[63 206] A selected image bank will require a smaller number of cases and observers to detect a small difference in performance but will have lower external validity.[63] Generalisability will be increased in a study that more closely replicates routine care, a heterogeneous mix of cases, but will require more cases and/or observers to identify small differences in observer performance.[63 183 206]

Another important consideration is the proportion of normal and abnormal cases required to ensure that the study is appropriately powered to detect a difference between observers or modalities without a prohibitively large number of cases and/or observers.[256] The type of observers also plays a role in determining the external validity of a diagnostic accuracy study.[183 355] Many of the Level 2 efficacy-diagnostic accuracy studies recruit highly specialised observers to interpret the index test.[183, 355] This is a recognised limitation when trying to apply the results to clinical practice as many of these diagnostic investigations demonstrate reduced performance when interpreted by average practitioners.[242] Nevertheless, the use of expert practitioners in the early development and assessment of imaging modalities plays a role, for if there is no improvement in the performance of these specialists there is unlikely to be when applied in routine care.[183] There is however, a paucity of literature that examines the performance of new radiology investigations in routine practice.

12.6 Non-Inferiority Approach

In an effort to improve the design, conduct and reporting of clinical trials so that they provide robust evidence to support practice, a set of suggested minimum standards was constructed by the Consolidated Standards of Reporting Trials (CONSORT) group who developed guidance and checklists,[356] similar to the STARD group,[47] for the various clinical trial designs. The checklist, adopted by the medical publishing community as best practice, was designed to improve the reporting of clinical trials. Recognition of biases due to study population or study design could help prevent flawed conclusions from being drawn and therefore inappropriately influencing practice. The CONSORT guidance can also be used to ensure that important determinants of bias, such as participant selection, exact nature of the intervention and statistical methodology, can be considered and addressed during study conception and design.[356] Although advocated as examples of best practice, uptake has not been universal.[357-359]

One significant difference in the current study is that the majority of medical research, both diagnostic and therapeutic, utilises a superiority approach; is treatment X more effective than placebo, does test Y have increased accuracy when compared to Z?[48] The aim of this study was not to investigate if the new test (reporting radiographers) were more accurate than the existing test (consultant radiologists), as radiographer reporting is always complimentary and integrated into a radiology service and is never intended to be a direct replacement for consultant radiologists.[44 168] The aim of the current study was to investigate if the performance of the reporting radiographers was no less effective, no less reliable in other words “non-inferior” to consultant radiologists’ reporting.

Non-inferiority studies have specific design considerations and statistical implications.[48 289 360] Standard superiority trials require a smaller sample size to accurately accept or reject the null hypothesis.[48 361] It is incorrect to assume equivalence or non-inferiority just because the null hypothesis has not been rejected.[48 238 241 289 360] These specific methodological issues, which reverse the traditional alternate and null hypotheses, have been addressed by an extension to the CONSORT statement and have been incorporated into both part one and part two of the current study.[48] These considerations were applied to the hypothesis statements below (Chapter 3.2). This has ensured that the study is adequately powered to detect a statistically significant difference, if one exists, between the diagnostic accuracy of chest X-ray interpretation of radiographers and radiologists and the influence that their reports have on clinicians' diagnostic thinking. While it may seem counterintuitive to have the null hypothesis stating that one treatment, intervention or group of reporting practitioners is superior to another, this methodology and statistical approach will facilitate robust and valid results.

12.7 Methods

The hierarchy of efficacy (Chapter 2.3), employed as the conceptual framework, identifies several paradigms which can be used to investigate each level. Utilising a staged approach, Part one of the study examined the accuracy with which qualified reporting radiographers interpreted a bank of chest x-rays in a controlled setting. Joint guidance published by the Royal College of Radiologists and Society and College of Radiographers reiterates that any radiographer who wishes to extend their practice to include image interpretation must perform at a level comparable to a consultant radiologist.[44] Therefore, in the current study, the performance of a group of consultant radiologists who interpreted the same image bank was the baseline measure.[48 289 360]

The written radiology report is the primary output of most radiology investigations and the vehicle for communicating with clinicians.[362] The role of the report is to communicate in an effective manner the nature, location and significance of the radiological findings, not to just describe the findings and assume that the clinician recognises the significance.[352] Part two of the study will investigate the role that radiographer and radiologist chest x-ray reports play in the diagnostic decision-making of clinicians; using the rationale that you must first change clinician diagnosis/confidence in order to instigate a change in management.[57 184 363]

The limitation of using diagnostic impact (level 3) as an intermediate outcome measure was recognised by Mackenzie and Dixon, who noted that "Diagnosis is not an end in itself; only a mental resting place for prognostic considerations and therapeutic decisions"(p.515).[61] Nevertheless, this step provides valuable information. If the investigation does not prompt a clinician to change their diagnosis or confidence in a diagnosis, there will be, by definition, no improvement in patient outcome but taking next step to show contribution of imaging to improved patient outcome is difficult.[133 184]

Several methodologies are available to examine this question, with distinct outcome measures. The methodologies chosen for each part were:

- Part One: Diagnostic Accuracy = alternative free-response receiver operator characteristic curves[82 183 242 244 338]
- Part Two: Diagnostic Impact = influence on diagnostic decision-making[62 148 164 233]

The evidence base for diagnosis had lagged behind therapeutic assessments, both in volume and quality.[59 133 364] Several large meta-analyses reviewed the methodological quality of diagnostic accuracy studies. These concluded that poor study design and poor reporting of key details, such as recruitment and application of the reference standard, meant that clinicians were unable to identify bias within these studies.[365 366] This is a challenge for clinicians who are required to assess if a diagnostic test would perform at a comparable level in routine practice, as the reported diagnostic accuracy within some research can be elevated due to bias.[367]

In order to increase the volume of robust research evidence to underpin practice, several statements which outline best practice have been published by eminent research groups for different research designs. These tools aid in the construction of solid methodologies, assist in the recognition and minimisation of biases, enable systematic reporting within the literature and contribute to consistency between studies. The STAndards for Reporting of Diagnostic accuracy (STARD) was published in 2003 as a 25 item checklist to facilitate high quality diagnostic accuracy research and improve study reporting through a structured and methodical approach.[46] This was updated in 2015 and consists of a 30-item checklist to be considered when performing studies that examine the accuracy of investigations.[47] This comprehensive list, ranging from title structure to presentation of results, has acted as a tool for study design and manuscript preparation. Recent work has suggested that, even in high impact journals, uptake has been heterogeneous and incomplete.[358 364] The current study has considered the 30 items as fundamental to a robust assessment of diagnostic accuracy of radiographer chest x-ray reporting and each has been addressed in the diagnostic accuracy methodology and methods.

Appendix 13 Electronic Records Search Terms

Examination Type = CR, DR

Examination Type = Chest

Patient Age is > 16 years

Patient Location = Accident & Emergency, Out-Patient, In-Patient

Report Status = Finalised

Report Completion time is 01/04/2011 < time < 31/03/2012

Appendix 14 <u>Advertisement for reporting radiographers in Synergy:Imaging and Therapy Practice and Synergy News (radiography professional newsletter)</u>

Calling all chest reporting radiographers!

We are conducting a study which will examine the accuracy of chest x-ray interpretation by qualified reporting radiographers.

If you

- Are a qualified chest x-ray reporting radiographer
- Currently undertake chest x-ray reporting sessions in clinical practice
- Working in the NHS

And:

- Can spare one day
- Are able to travel to (xxxx) London

What are you waiting for?!

To register your interest or for further information contact xxxx, chief investigator, on xxxx or xxxx.

Appendix 15 Data collection template for reference standard diagnosis – Expert Chest Radiologist

Reference Standard Report Proforma

Chest X-ray Image Bank *n*

Reviewer: CCR1/2

There are *n* chest x-rays for interpretation in this image bank which have been randomly selected from hospital based patients at a London acute district general hospital (DGH).

Section 1 contains the patient demographics, including age, gender and source of referral (Accident & Emergency, In-Patient or Out-Patient). If any previous chest x-rays are available this will be indicated, and will be located in the patients' image folder.

Section 2 provides the clinical information and history given to the reporting practitioner at the time of initial imaging request.

Section 3 is to be completed by placing an **'x'** in the corresponding box indicating your interpretation of the chest x-ray. A list of abnormalities not considered to be clinically relevant is included (Appendix 1). For the purposed of this study, any images containing only these findings should be interpreted as **normal**.

Section 4 is a free text field which should contain your report, containing a description of the salient features, diagnosis (if appropriate) and any specific recommendations. Please localise all abnormalities using the zonal criteria (Appendix 2).

Section 5 asks for a list of all abnormalities (if any) and their location on the image, using the predefined zonal localisation criteria (Appendix 2).

Section 6 requires that you allocate a disease category for each abnormal case.

Section 7 requires an **'x'** to be placed in the box corresponding to the level of conspicuity of the lesion described in an abnormal report. Please leave blank if in your opinion the radiograph is normal.

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

Case Number: 0001	**Case Reference Number:** Anonmale1946		
Section 1 - Patient Demographics			
Age: xx years	Gender: Fe/Male	Referral Source: AE/IP/OP	Previous Chest X-rays: Yes/No
Section 2 - Clinical History			
(as provided by clinician at the time of initial request)			
Section 3 – Confidence in Interpretation (please mark one box below with an **x**)			

0	1
Normal	**Abnormal**

Section 4 - Report

(free text of salient findings, diagnosis and recommendations)

Section 5 – List of Abnormalities

*(please **list** and **localise** all abnormalities eg. RUZ consolidation, L apical fibrosis etc)*

Section 6 – Disease Category

(please mark one box with a **'x'** if case is abnormal)

Infection	**Cardiac/Pulm Oedema**	**Malignancy**	**Other**

Section 7 – Conspicuity (please mark one box with a **'x'**)

1	2	3	4
Very Subtle	**Subtle**	**Obvious**	**Very Obvious**

appendix 1 – Reporting Guidance

(This will be based on the work of Robinson *et al*.[188] and agreed after discussion with the expert chest consultant radiologists prior to the study commencing)

For the purposes of this study, the following are to be considered **NORMAL**:

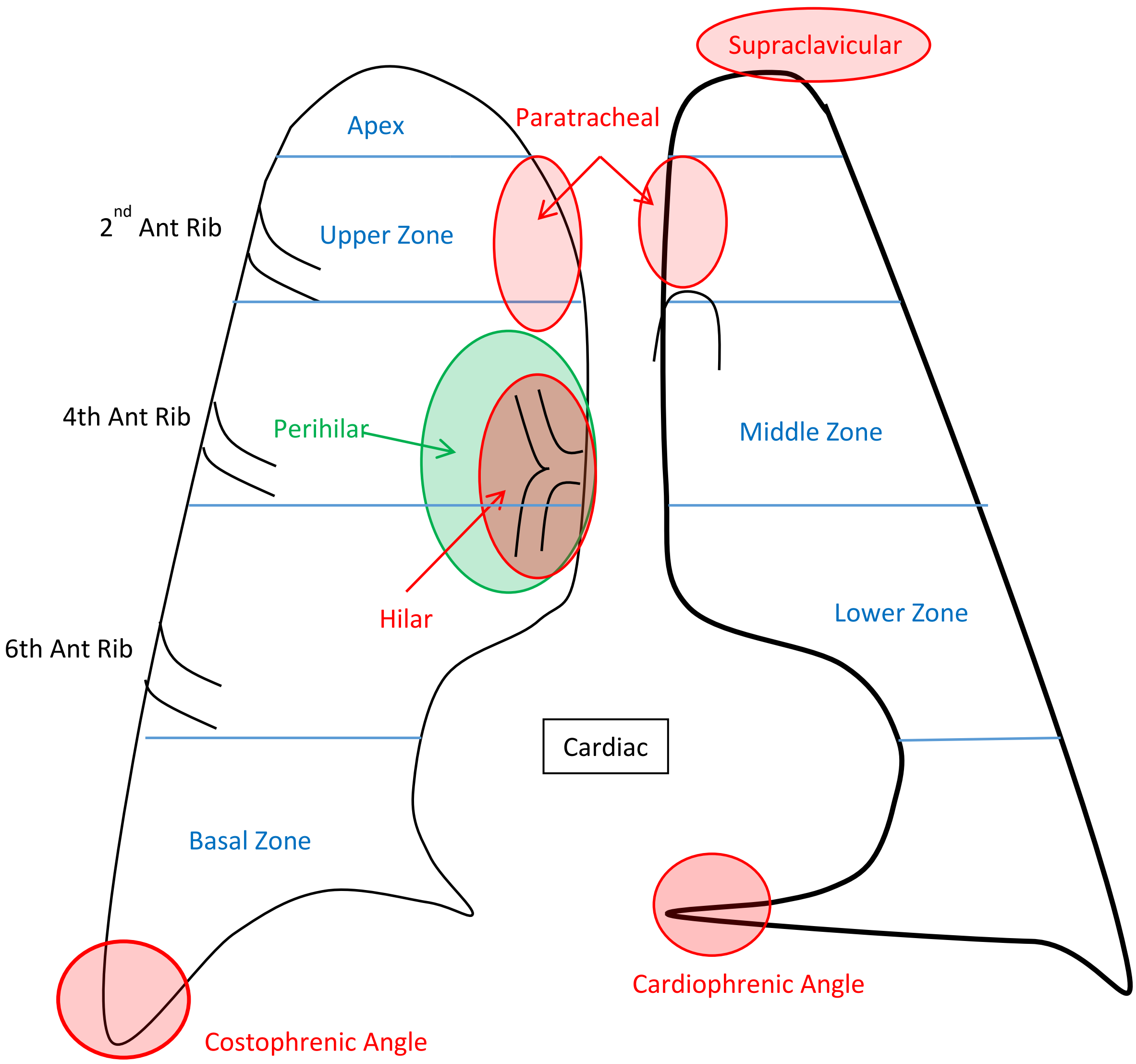
- Congenital/anatomical variants
- Small calcified foci
- Old fractures
- Previous surgery
- Hiatus hernia

For the purposes of this study, the following are to be considered **ABNORMAL**:

- Pleural fluid
- Pneumothorax
- Consolidation
- Non-calcified nodules
- Cardiac enlargement
- Mediastinal widening
- Recent fracture
- Foreign body

Appendix 2 – Zonal Localisation Criteria

For the purposes of this study please use the following zonal criteria when localising abnormalities.



Appendix 16 Arbiter agreement proforma – reference standard

Arbiter Proforma

Reference Standard

Chest X-ray Image Bank *n*

Reviewer: *A1/2*

There are *n* chest x-ray reports for comparison, derived from an image bank which has been randomly selected from hospital based patients at a London acute district general hospital (DGH).

Section 1 provides the reports for each case.

Section 2 asks for your opinion regarding agreement between the abnormality lists and select either **agree** or **disagree** by placing a **'x'** in the corresponding box.

Please use the acceptance radius as outlined in Appendix 1.

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

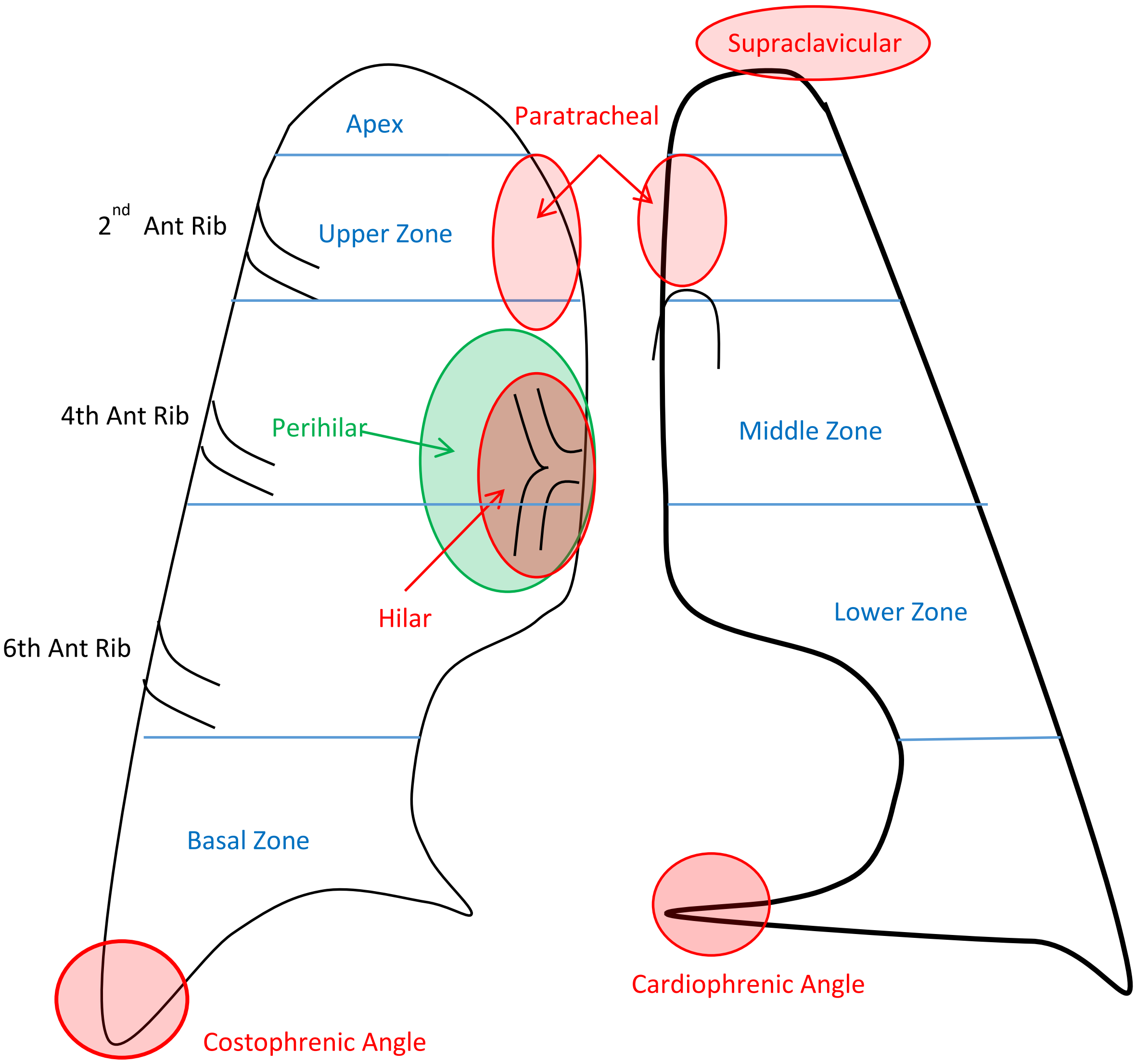
Abnormality List Number: 001	
Section 1 – Abnormality Lists	
List A	List B
Section 2 – Report Agreement	

Please mark a box with an **'x'**

0	**1**
Agree	Disagree

appendix 1 – Zonal Localisation Criteria[250]

For agreement to be reached, the abnormality described in the lists above must be contained with the same zonal region (acceptance radius). For the purposes of this study please use the following zonal criteria when localising abnormalities.



Appendix 17 Participant (consultant radiologist/reporting radiographer) chest X-ray interpretation data collection proforma

Participant Report Proforma

Chest X-ray Image Bank A & B

Reviewer: R*n*

There are 106 chest radiographs for interpretation in his image bank which have been randomly selected from hospital based patients at a London acute district general hospital (DGH).

Section 1 contains the patient demographics, including age, gender and source of referral (Accident & Emergency, In-Patient or Out-Patient). If any previous chest x-rays are available this will be indicated, and will be located in the patients imaging folder.

Section 2 provides the clinical information and history given to the reporting practitioner at the time of initial imaging request.

All information (Sections 3 & 4) are to be completed on the online form.

Section 3 is to be completed by selecting the corresponding box indicating your interpretation of the chest x-ray. A list of abnormalities not considered to be clinically relevant is included (Appendix 1). For the purposed of this study, any images containing only these findings should be interpreted as **normal**.

Section 4 is a free text field which should contain your report, containing a description of the salient features, diagnosis (if appropriate) and any specific recommendations. Please localise all abnormalities using the zonal criteria (Appendix 2). All abnormalities must be assigned a confidence score by inserting a rating after the free text description (1 – 4), according to the following scale:

1 = uncertain
2 = possibly abnormal
3 = probably abnormal
4 = definitely abnormal

For example "*A 2cm right upper zone nodule is suspicious for a malignant lesion (3).*"
Normal images do not require a confidence rating.

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

<table>
<tr><td colspan="2">Case Number: 013</td><td colspan="2">Case Reference Number: 4120020A</td></tr>
<tr><td colspan="4">Section 1 - Patient Demographics
Request Number: 11000099A</td></tr>
<tr><td><u>Age:</u> 73 years</td><td><u>Gender:</u> Male</td><td><u>Referral Source:</u> OP</td><td><u>Previous:</u> Yes</td></tr>
<tr><td colspan="4">Section 2 - Clinical History</td></tr>
<tr><td colspan="4">(as provided by clinician at the time of initial request)
Previous lung ca surgery-?recurrence
smoker</td></tr>
</table>

Appendix 1 – Reporting Guidance

For the purposes of this study, the following are to be considered **NORMAL**:

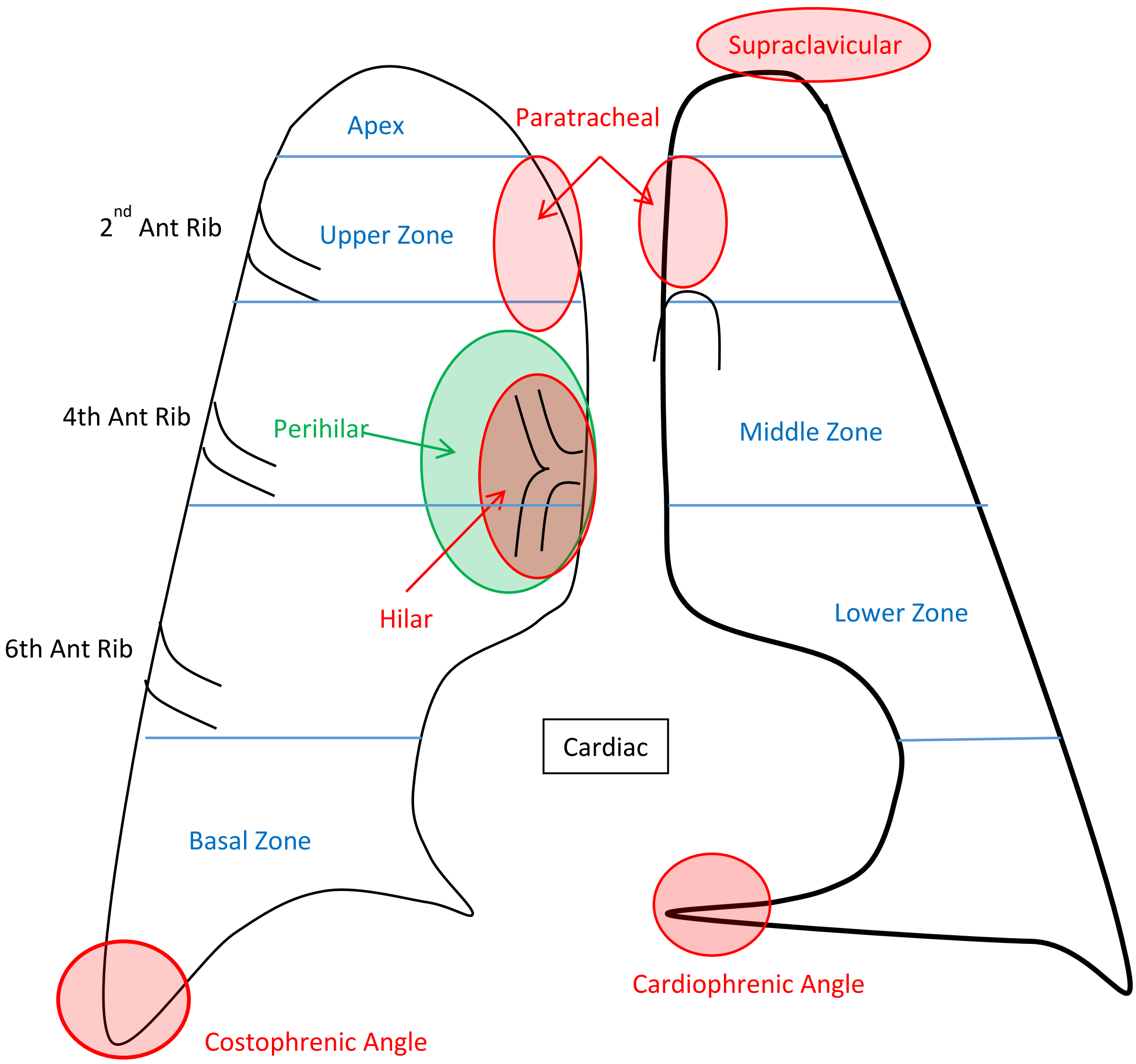
- Congenital/anatomical variants
- Small calcified foci
- Old fractures
- Previous surgery
- Hiatus hernia

For the purposes of this study, the following are to be considered **ABNORMAL**:

- Pleural fluid
- Pneumothorax
- Consolidation
- Non-calcified nodules
- Cardiac enlargement
- Mediastinal widening
- Recent fracture
- Foreign body

Appendix 2 – Zonal Localisation Criteria

For the purposes of this study please use the following zonal criteria when localising abnormalities.



Appendix 18 Arbiter proforma – participant observers (consultant radiologists and reporting radiographers) for diagnostic accuracy study (Part 1)

Arbiter Proforma
Participant Reports
Chest X-ray Image Bank *n*
Reviewer: *A1/2*

There are *n* chest x-rays reports which have been produced from an image bank randomly selected from hospital based patients at a London acute district general hospital (DGH).

Section 1 provides the reports for each case.

Section 2 asks you to extract all abnormalities from the participant report with the corresponding confidence score. Using the reference standard diagnosis list, assign either a *lesion localisation* or *non-lesion localisation* rating for each abnormality.

An abnormality identified in the participant report will be deemed to be a *lesion localisation* (LL) event **if**:

- the abnormality is within the acceptance radius (Appendix 1)
- the correct diagnosis is made

An abnormality in the participant report will be deemed a *non-lesion localisation* (NLL) event **if**:

- the abnormality is outside the acceptance radius (Appendix 1)
  and/or
- the incorrect diagnosis is made

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

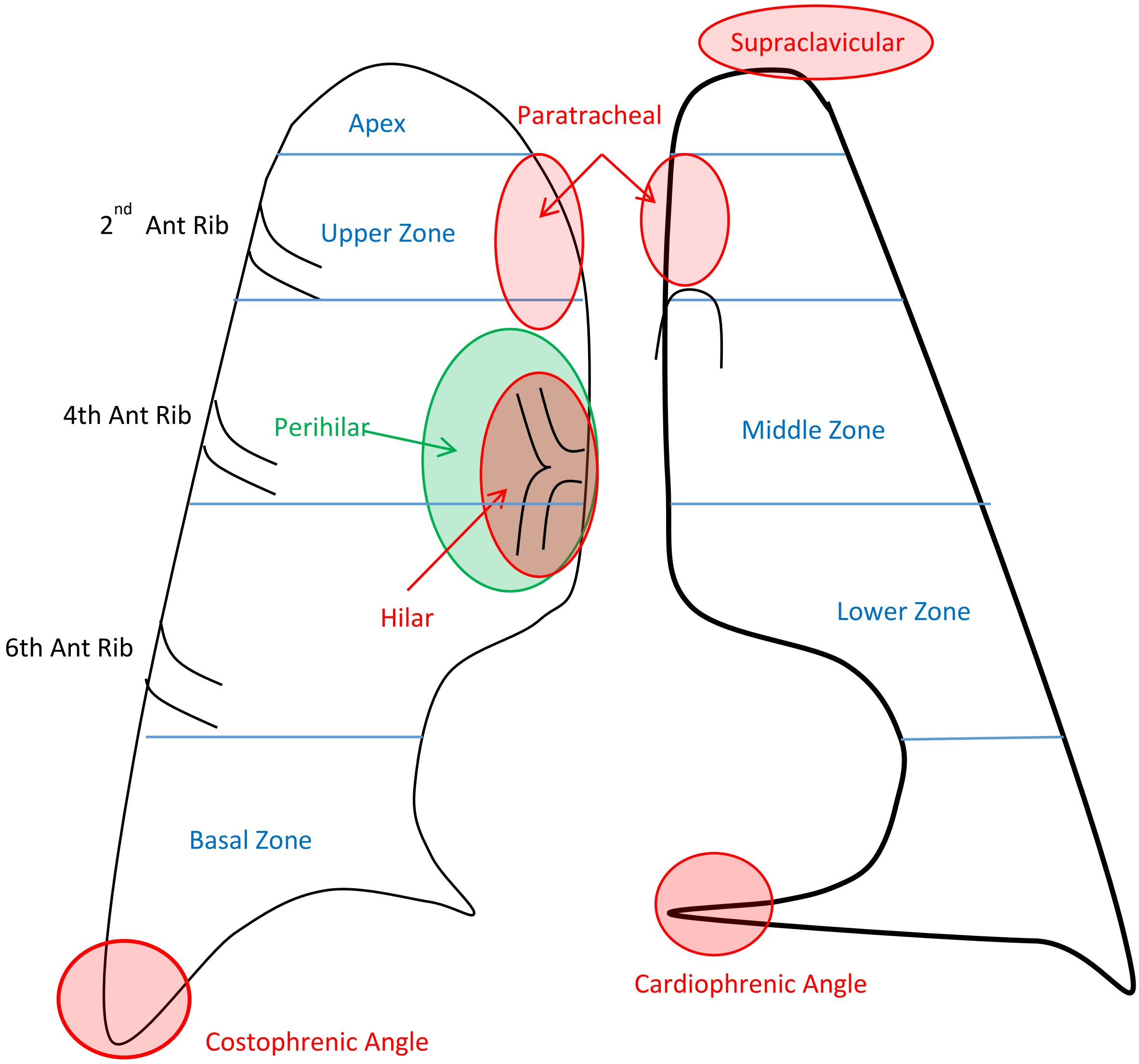
Report Pair Number: 001	
Section 1 – Radiology Reports/Abnormality Lists	
Reference Standard Abnormality List	Participant Report
Section 2 – Report Agreement	

Please complete the table below for the participant report provided

Abnormality	Rating	LL/NLL

appendix 1 – Zonal Localisation Criteria

For agreement to be reached, the abnormality described in the lists above must be contained with the same zonal region (acceptance radius). For the purposes of this study please use the following zonal criteria when localising abnormalities.



Appendix 19 JAFROC lesion master sheet (sample)

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Demographics				Other Exams			CCR Dx		Agreement		Conspicuity		Lesion 1		Lesion 2		Lesion 3		Lesion 4		Lesion 5		Lesion 6	
2	Study Ref	Age	Exam Date	Encounter Type	Prev CXR	F/U CXR	Clinical Report	CCR1	CCR2	A1	A2	CCR1	CCR2	Locat	Weight	Locat	Weight	Locat	Weight	Locat	Weight	Locat	Weight	Locat	Weight
3	13	73		2	1	1	4	4	0	1	1	3	1												
4	1	27		1	1	0	2	2	4	1	1	3	1	Heart		Spine									
5	2	49		0	0	0	0	0	0	1	1	0	0												
6	3	47		1	0	0	0	0	0	1	1	0	0												
7	4	83		0	1	0	4	4	4	1	1	4	4	LMZ		RMZ		R Diaph							
8	30	66		1	1	0	4	4	4	1	1	3	2	LLZ		RLZ									
9	8	77		0	1	1	0	0	0	1	1	0	0												
10	10	36		2	1	0	0	0	0	1	1	0	0												
11	15	20		2	1	0	0	0	0	1	1	0	0												
12	17	69		2	1	0	0	0	0	1	1	0	0												
13	20	36		0	0	0	0	0	0	1	1	0	0												
14	21	68		1	1	1	1	2	2	1	1	3		LMZ		RMZ		LLZ		RLZ		L CP			
15	22	52		2	1	0	0	0	0	1	1	2	0												
16	23	33		1	0	1	1	1	2	1	1	3	3	RLZ		R CP									
17	24	78		0	1	1	2	4	0	1	1	0	0												
18	28	60		0	1	1	0	0	0	1	1	0	0												
19	45	19		2	1	1	1	1	1	1	1	3	3	L Apex		R Apex									
20	29	53		1	0	1	4	4	4	1	1	3	2	RLZ											
21	31	86		0	1	1	2	2	2	1	1	3	2	LMZ		RMZ									
22	43	40		1	1	0	1	4	4	1	1	3	2	L CP											
23	39	50		0	1	1	3	1	3	1	1	3	2	LUZ		L CP									
24	40	80		1	1	1	3	2	2	1	1	3	3	LMZ		RMZ/RLZ		L CP		R CP					
25	41	16		1	1	1	0	0	0	1	1	0	0												
26	42	52		0	0	1	0	2	2	1	1	2	3	LUZ/LMZ		RMZ									
27	46	54		1	0	1	0	0	0	1	1	0	0												
28	47	74		1	1	1	2	2	2	1	1	3	3	LLZ		LMZ		RMZ							
29	52	61		0	1	1	2	4	4	1	1	3	3	RLZ (NGT)		LLZ									
30	53	68		0	1	1	2	2	2	1	1	3	3	Heart		RLZ		R CP							
31	54	44		2	0	0	0	0	0	1	1	0	0												
32	57	42		0	0	1	0	0	0	1	1	0	0												
33	58	34		1	1	1	1	0	0	1	1	0	0												
34	59	36		0	0	0	0	0	0	1	1	0	0												
35	60	62		0	1	1	4	4	4	1	1	3	3	L Apex		R Apex		L Lung		R Lung					
36	61	84		1	0	1	0	0	0	1	1	0	0												
37	62	90		1	1	1	1	1	3	1	1	3		RLZ (NGT)		RUZ		RLZ		LLZ		L CP		R CP	
38	64	76		0	1	1	3	3	3	1	1	3	3	LUZ		LLZ		RLZ							
39	66	27		0	0	0	0	0	0	1	1	0	0												
40	68	56		0	1	1	0	0	0	1	1	0	0												
41	69	65		0	1	0	0	0	0	1	1	0	0												
42	70	56		0	1	0	0	0	0	1	1	0	0												
43	71	52		0	1	1	4	4	4	1	1	3	3	L Lung		R Lung									
44	74	46		0	0	1	0	0	0	1	1	0	0												
45	76	41		0	0	0	0	0	0	1	1	0	0												
46	79	60		0	1	0	2	2	2	1	1	3	2	LMZ/LLZ		RMZ/RLZ									
47	81	71		0	1	0	4	2	2	1	1	3	1	LUZ		RUZ									

Appendix 20 <u>Sample of JAFROC lesion localisation data collation spreadsheet</u>

True Positive Events (section)

Study Ref	Locat	AbN	R2	R3	R4	R5	R6	R18	R7	R21	R1	R10	R9	R16	R17	R19	R8	R12	R13	R22	R24	R20	R23
			1	1	1	2	1	2	2	2	1	2	2	1	1	2	1	1	2	1	1	2	2
13			xx	xx	xx	0	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
1	1 Heart	Cardiomegaly	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	MISS	4	0	4	4	4
	2 Spine	AVN	4	0	0	0	0	0	4	0	0	0	0	0	0	0	0	MISS	0	0	0	0	0
2			xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	0	xx	0	xx	xx	MISS	xx	xx	xx	xx	xx
3			xx	xx	xx	xx	xx	0	xx	xx	xx	xx	xx	xx	xx	xx	xx	0	xx	xx	xx	xx	xx
4	1 LMZ	Pleural Calc	4	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	0	4	4	4	4
	2 RMZ	Pleural Calc	4	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	0	4	4	4	4
	3 R Diaph	Pleural Calc	4	3	4	4	4	4	4	4	4	4	3	4	4	2	4	4	0	4	0	4	4
30	1 LLZ	Fibrosis	4	4	0	0	4	4	0	2	0	4	0	4	3	4	4	4	2	4	3	0	4
	2 RLZ	Fibrosis	0	4	0	0	4	0	4	2	0	4	0	4	0	2	4	0	2	0	3	0	4
8			0	xx	xx	0	xx	xx	xx	0	xx	xx	0	0	0	0	0	xx	xx	0	xx	xx	xx
10			xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
15			xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
17			xx	xx	0	0	xx	xx	0	xx	xx	xx	0	0	xx	0	xx	xx	xx	0	xx	0	0
20			xx	0	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	0	xx	xx	xx	xx	xx	xx	xx
21	1 LMZ	Interstitial oec	4	4	0	0	0	0	0	4	0	4	0	0	4	4	0	4	0	0	3	0	0
	2 RMZ	Interstitial oec	4	4	0	0	0	0	0	4	0	4	0	4	4	4	0	4	0	0	3	0	0
	3 LLZ	Pulm Oedema	0	4	4	4	4	4	3	4	4	0	0	0	4	4	4	0	0	4	4	0	4
	4 RLZ	Pulm Oedema	4	4	4	4	4	4	0	4	4	0	4	4	4	4	4	4	4	4	4	3	4
	5 L CP	Effusion	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
22			xx	xx	xx	0	xx	xx	xx	xx	0	xx	xx	xx	xx	xx	0	xx	xx	xx	xx	xx	xx
23	1 RLZ	Consolidation	4	3	4	4	4	4	4	0	4	0	4	4	4	4	4	4	4	4	4	0	0
	2 R CP	Effusion	4	3	4	0	0	0	4	0	0	0	0	4	4	3	4	0	0	4	3	0	3
24			xx	0	xx	0	xx	xx	0	xx	xx	0	xx	xx	0	xx	0	xx	xx	0	0	xx	0
28			xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
45	1 L Apex	Fibrosis	4	0	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	0	4
	2 R Apex	Fibrosis	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
29	1 RLZ	Rib #	4	4	4	0	4	0	4	4	4	4	4	4	0	4	4	4	4	4	4	0	4
31	1 LMZ	Interstitial Oec	4	4	4	0	4	0	0	0	4	4	0	4	4	4	4	0	0	3	0	0	4
	2 RMZ	Interstitial Oec	4	4	4	0	4	0	0	0	4	4	3	4	4	4	4	0	0	3	0	0	4
43	1 L CP	Effusion	2	0	1	2	4	2	4	4	3	4	0	4	4	4	4	4	4	0	3	0	0

False Positive Events (section)

Study Ref	R2	R3	R4	R5	R6	R18	R7	R21	R1	R10	R9	R16	R17	R19	R8	R12	R13	R22	R24	R20	R23
13				3																	
1		3	4	2	3	2		4		4	4	4	4	4	4		2		4		2
		3	4		3			4		4	4		4	4	4				4		2
2											1		4								
													4								
3						2															
4		4	4	3	2	2	4		4		3	4	4	3	4		4	4	4		2
		2	3									4									
30				4		2	4		4	4	2	4			4		3		4	2	4
				4					4	4	2	4			4					2	4
				4							2									1	4
				2																2	
																				2	
8	2			4				4			1	4	4	2	4			4			
	2			4				4				4	4	2	4			4			
				4				4													
								4													
10																					
15																					
17			3	3			3				1	4		2				2		3	3
																				3	3
20														1							
21				4		3				4	4	4		4	4				3	4	
				4							4			4	4					4	
															4						
22				4					3						2						
				4																	
23	4			4	4	3	3	4	4	1	4		4	4	4	4	4	4	4	4	4
				4					4		4		4							3	4
				4					4											3	3
24		3		4			3			1			3		4			3	2		2
													3		4				2		
28																					
45		2			4					4		4		4			3			2	

Appendix 21 Clinician recruitment email

Dear Colleague,

We are conducting a study which will examine the influence that chest x-ray reports have on clinicians' diagnostic thinking.

You will be required to review a series of short case scenarios and to arrive at a diagnosis, both with and without a chest x-ray report. The scenarios will be given to you in small batches to be completed at your convenience.

To register your interest or for further information please contact xxxx, chief investigator, at xxxx or xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

 Clinician pre-chest X-ray diagnosis proforma

Clinician Report Proforma

Diagnostic Case Summaries Bank *n*
Reviewer: C*n*

There are *n* case summaries in this bank which have been randomly selected from hospital based patients at a London acute district general hospital (DGH) who have been referred for a chest x-ray as part of their management.

In the context of each individual case, please complete Sections 3 and 4 overleaf.

Section 1 contains the patient demographics, including age, gender and source of referral (Accident & Emergency, In-Patient or Out-Patient). If any previous chest x-rays have been obtained this will be indicated.

Section 2 provides the summary for each case. This will include patient history, symptoms, clinical findings and the results of laboratory investigations available prior to the initial chest x-ray request being made.

Section 3 requires you to formulate your most likely diagnosis from the information provided and to select one option only by placing a **'x'** in the corresponding box from the list provided. If this list does not contain your most likely diagnosis then please select the 'other' option. You should also consider your confidence in the 'most likely' diagnosis and select this by placing a numerical score in the box between 0 and 100, where:

0 = very unlikely
50 = uncertain
100 = certain

Section 4 requires you to consider your 'most important' diagnosis for the presented case summary. This is the most serious diagnosis that you would want to exclude in this patient and one option should be selected from the list provided. You should also indicate your confidence in the 'most important' diagnosis and select this by placing a numerical score in the box between 0 and 100, where:

0 = very unlikely
50 = uncertain
100 = certain

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on 0208 510 7848 or 0208 510 7105.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

Case Number: 0001	
Section 1 - Patient Demographics	
Age: xx years	Gender: Fe/Male
Referral Source: AE/IP/OP	Previous Chest X-rays: Yes/No
Section 2 – Case Summary	
(case summary provided by a Physician following review of the notes, to include all pertinent information available up to the time of initial chest x-ray request)	

Section 3 – Most Likely Diagnosis

Most Likely Diagnosis
(please mark one box below with an **x**)

Diagnosis			6	Pneumothorax	
1	Normal		7	Perforation	
2	Infection		8	Malignancy	
3	Cardiac/Pulm Oedema		9	TB	
4	Pleural Effusion		10	Fibrosis	
5	COPD		11	Other	

Confidence in Most Likely Diagnosis
(please provide your numerical answer below)

Confidence (0-100)	

Section 4 – Most Important Diagnosis

Most Important Diagnosis
(please mark one box below with an **x**)

Diagnosis			6	Pneumothorax	
1	Normal		7	Perforation	
2	Infection		8	Malignancy	
3	Cardiac/Pulm Oedema		9	TB	
4	Pleural Effusion		10	Fibrosis	
5	COPD		11	Other	

Confidence in Most Important Diagnosis
(please provide your numerical answer below)

Confidence (0-100)	

Appendix 23 Clinician post-chest X-ray diagnosis proforma

Clinician Report Proforma
Diagnostic Case Summaries Bank *n*
Reviewer: C*n*

There are *n* case summaries in this bank which have been randomly selected from hospital based patients at a London acute district general hospital (DGH) who have been referred for a chest radiograph as part of their management.

In the context of each individual case, please complete Sections 4 and 5 overleaf.

Section 1 contains the patient demographics, including age, gender and source of referral (Accident & Emergency, In-Patient or Out-Patient). If any previous chest x-rays have been obtained this will be indicated.

Section 2 provides the summary for each case. This will include patient history, symptoms, clinical findings and the results of laboratory investigations available prior to the initial chest x-ray request being made.

Section 3 provides a radiology report for the chest x-ray that was requested by the treating clinician.

Section 4 requires you to formulate your most likely diagnosis from the information provided and to select one option only by placing a **'x'** in the corresponding box from the list provided. If this list does not contain your most likely diagnosis then please select the 'other' option. You should also consider your confidence in the 'most likely' diagnosis and select this by placing a numerical score in the box between 0 and 100, where:

0 = very unlikely
50 = uncertain
100 = certain

Section 4 requires you to consider your 'most important' diagnosis for the presented case summary. This is the most serious diagnosis that you would want to exclude in this patient and one option should be selected from the list provided. You should also indicate your confidence in the 'most important' diagnosis and select this by placing a numerical score in the box between 0 and 100, where:

0 = very unlikely
50 = uncertain
100 = certain

Many thanks for your participation in this study. If there are any queries please do not hesitate to contact the researcher on xxxx or xxxx.

This study has been given formal ethics approval by xxxx
Local R&D approval reference xxxx

Case Number: 0001			**Report Reference:** vwxyz
Section 1 - Patient Demographics			
Age: xx years	Gender: Fe/Male	Referral Source: AE/IP/OP	Previous Chest X-rays: Yes/No

Section 2 – Case Summary

(case summary provided by a Physician following review of the notes, to include all pertinent information available up to the time of initial chest x-ray request)

Section 3 – Radiology Report

(report produced by one of the reporting practitioners in the diagnostic accuracy arm of the study, anonymised for reporting source [radiologist/radiographer])

Section 4 – Most Likely Diagnosis

Most Likely Diagnosis
(please mark one box below with an x)

Diagnosis		Pneumothorax	
Normal		Perforation	
Infection		Malignancy	
Cardiac/Pulm Oedema		TB	
Pleural Effusion		Fibrosis	
COPD		Other	

Confidence in Most Likely Diagnosis
(please provide your numerical answer below)

Confidence (0-100)	

Section 5 – Most Important Diagnosis
This is the most serious diagnosis you would want to exclude

Most Important Diagnosis
(please mark one box below with an x)

Diagnosis		**Diagnosis**	
Infection		Perforation	
Cardiac/Pulm Oedema		Malignancy	
Pleural Effusion		TB	
COPD		Fibrosis	
Pneumothorax		Other	

Confidence in Most Important Diagnosis
(please provide your numerical answer below)

Confidence (0-100)	

Appendix 24 <u>Clinicians diagnoses and diagnostic confidence data collation sheet (sample)</u>

Reporting_I	Rep_Pract	Most Likely	Probability	Likely Cat	Likely Comp	Novel Likel	New Correct	effect of CX	Corrected C	Sig Likely C	Sig Likely C	Most Seriou	Probability	Serious Cat	Serious Con	Novel Serio	Novel Serio	Effect CXR S	Corrected C	Sig Serious	Sig Serious	Likely or Se	New Likely	Sig Serious	Likely/Seric	Sig Likely/S	Both Likely/
8	1	1	90	4	0	0	0	4	-30	0	0	1	80	4	0	0	0	4	-20	0	0	0	0	0	0	0	0
8	1	1	95	5	0	1	0	4	-5	0	0	1	95	5	0	0	0	4	-75	0	0	0	0	0	0	0	0
8	1	1	50	3	0	1	0	1	-140	0	0	1	50	3	0	0	0	4	-40	0	0	0	0	0	0	0	0
8	1	0	60	3	0	0	0	4	30	1	0	99	10	2	3	0	0	3	5	0	0	1	0	1	0	0	0
8	1	99	95	5	0	0	0	4	0	0	0	99	5	1	0	1	0	1	-15	0	0	0	0	0	0	0	0
8	1	1	5	1	0	1	0	4	85	1	0	1	5	1	0	1	0	4	45	1	0	0	0	1	1	0	0
8	1	0	80	4	0	1	0	1	-140	0	0	7	10	2	0	0	0	4	30	1	0	0	0	1	0	0	0
8	1	0	99	5	3	0	0	3	0	0	0	1	1	1	0	1	0	4	0	0	0	1	0	0	0	0	0
8	1	0	95	5	3	0	0	3	-3	0	0	7	3	1	0	0	0	4	-2	0	0	1	0	0	0	0	0
8	1	0	95	5	3	0	0	3	15	0	0	99	5	1	0	0	0	4	15	0	0	1	0	0	0	0	0
8	1	2	90	4	3	0	0	3	30	1	0	7	1	1	0	0	0	4	9	0	0	1	0	1	0	0	0
8	1	0	80	4	0	1	0	1	-130	0	0	1	10	2	0	0	0	4	10	0	0	0	0	0	0	0	0
8	1	2	60	3	3	1	1	2	159	1	0	2	60	3	3	0	0	3	59	1	0	1	1	1	1	0	0
8	1	1	100	5	0	1	0	4	-60	0	0	1	100	5	0	0	0	4	-80	0	0	0	0	0	0	0	0
8	1	8	60	3	0	1	0	4	0	0	0	8	60	3	0	1	0	4	-30	0	0	0	0	0	0	0	0
8	1	0	99	5	0	0	0	4	0	0	0	0	99	5	0	1	0	4	-98	0	0	0	0	0	0	0	0
8	1	1	90	4	0	1	0	4	9	0	0	1	90	4	0	1	0	4	-89	0	0	0	0	0	0	0	0
8	1	2	90	4	3	1	1	2	150	1	0	2	90	4	3	0	0	3	60	1	0	1	1	1	1	0	0
8	1	7	98	5	0	1	0	4	-28	0	0	7	98	5	0	0	0	4	-97	0	0	0	0	0	0	0	0
8	1	2	99	5	3	0	0	3	29	1	0	2	99	5	3	0	0	3	29	1	0	1	0	1	1	0	0
8	1	0	60	3	0	1	0	1	-155	0	0	1	20	2	3	0	0	3	-75	0	0	1	0	0	0	0	0
8	1	2	70	4	0	1	0	4	25	1	0	1	20	2	0	0	0	4	-15	0	0	0	0	1	0	0	0
8	1	3	99	5	0	1	0	4	-59	0	0	3	99	5	0	0	0	4	-79	0	0	0	0	0	0	0	0
8	1	8	100	5	0	0	0	4	0	0	0	8	100	5	0	0	0	4	-1	0	0	0	0	0	0	0	0
8	1	4	60	3	0	1	0	4	39	1	0	1	5	1	0	0	0	4	-4	0	0	0	0	1	0	0	0
8	1	2	70	4	0	1	0	1	-80	0	0	1	20	2	3	1	1	2	110	1	0	1	1	1	0	0	0
8	1	1	90	4	0	1	0	4	0	0	0	1	90	4	0	1	0	1	-100	0	0	0	0	0	0	0	0
8	1	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888
8	1	1	80	4	0	0	0	4	-10	0	0	1	80	4	0	0	0	4	-10	0	0	0	0	0	0	0	0
8	1	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888
8	1	99	90	4	0	1	0	4	-30	0	0	99	90	4	0	1	0	4	-60	0	0	0	0	0	0	0	0
8	1	0	95	5	0	0	0	4	-5	0	0	2	2	1	3	1	1	2	7	0	0	1	1	0	0	0	0
8	1	1	80	4	0	1	0	1	-130	0	0	1	80	4	0	0	0	4	-40	0	0	0	0	0	0	0	0
8	1	0	90	4	0	0	0	4	5	0	0	7	3	1	3	1	1	2	4	0	0	1	1	0	0	0	0
8	1	1	80	4	0	0	0	4	0	0	0	1	80	4	0	0	0	4	0	0	0	0	0	0	0	0	0
8	1	1	95	5	0	0	0	4	-25	0	0	7	3	1	3	1	1	2	93	1	1	1	1	1	0	1	0
8	1	0	70	4	3	0	0	3	10	0	0	99	20	2	0	1	0	4	0	0	0	1	0	0	0	0	0
8	1	99	90	4	0	0	0	4	-10	0	0	1	5	1	3	1	1	2	15	0	0	1	1	0	0	0	0
8	1	2	70	4	3	0	0	3	0	0	0	2	70	4	3	0	0	3	0	0	0	1	0	0	0	0	0
8	1	0	95	5	3	0	0	3	15	0	0	99	2	1	0	1	0	4	3	0	0	1	0	0	0	0	0
8	1	1	90	4	0	1	0	1	-180	0	0	1	90	4	0	0	0	4	-85	0	0	0	0	0	0	0	0
8	1	0	90	4	0	1	0	1	-185	0	0	2	5	1	0	1	0	4	0	0	0	0	0	0	0	0	0
8	1	99	80	4	0	0	0	4	-50	0	0	99	80	4	0	1	0	4	-75	0	0	0	0	0	0	0	0
8	1	1	90	4	0	1	0	4	-80	0	0	1	90	4	0	0	0	4	-80	0	0	0	0	0	0	0	0
8	1	99	100	5	0	1	0	4	-80	0	0	99	100	5	0	1	0	4	-95	0	0	0	0	0	0	0	0
8	1	3	100	5	0	1	0	1	-200	0	0	7	100	5	3	1	1	2	105	1	1	1	1	1	0	1	0
8	1	0	97	5	3	0	0	3	0	0	0	1	2	1	0	1	0	4	1	0	0	1	0	0	0	0	0
8	1	2	90	4	0	1	0	1	-180	0	0	99	5	1	0	1	0	4	0	0	0	0	0	0	0	0	0
8	1	99	80	4	0	1	0	1	-180	0	0	7	10	2	0	0	0	4	10	0	0	0	0	0	0	0	0
8	1	0	95	5	3	0	0	3	15	0	0	8	2	1	0	0	0	4	13	0	0	1	0	0	0	0	0
8	1	0	95	5	3	0	0	3	15	0	0	8	1	1	0	0	0	4	14	0	0	1	0	0	0	0	0
8	1	7	90	4	3	1	1	2	120	1	0	7	90	4	3	1	1	2	92	1	0	1	1	1	1	0	0
8	1	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888	888