An Evaluation of Radiographer Plain Radiograph Reporting

Stephen Brealey

Submitted for the degree of Doctor of Philosophy

in Health Sciences

Department of Health Sciences

University of York

2003

Abstract

Historically, reporting of images on radiographs or other recording media has been the domain of radiologists. A shortage of these professionals and changes in government policy over the last decade have resulted in relaxation of restrictions on who should report these images, providing radiographers with an opportunity to develop their reporting roles, as reflected by the increase in numbers reporting Accident and Emergency (A&E) radiographs from four Trusts in 1995 to 32 Trusts in 1999. In order to establish whether this increase in radiographer reporting is justified, the thesis aims to evaluate whether selectively trained radiographers should report A&E plain radiographs and also the potential for further extending their reporting role.

A systematic review provides evidence that selectively trained radiographers are able to report A&E radiographs to a high level of accuracy. There is also no evidence to suggest that radiologists of varying seniority report these radiographs more accurately than selectively trained radiographers.

Primary research from the thesis also provides evidence that there is no statistically or clinically significant difference in A&E radiograph reporting performance between consultant radiologists and selectively trained radiographers, nor in the subsequent clinical effects of their reports on clinicians' diagnoses, choice of management plans, and patient outcome. No obvious cost savings are obtained from substituting consultant radiologists with selectively trained radiographers in an A&E plain radiograph reporting role. No clear evidence indicates that consultant radiologists report GP plain radiographs significantly more accurately than selectively trained radiographers. Some of the findings suggest a more adverse effect on GPs' confidence in their diagnoses and management plans following incorrect reports by radiologists than those of the selectively trained radiographers, although this is not reflected in patient outcome.

The main conclusion of the thesis is that selectively trained radiographers can substitute radiologists for the reporting of A&E plain radiographs and X-ray departments should invest in this skill mix, if it can help meet local demands. Further research is needed before the same conclusion can be drawn about selectively trained radiographers reporting GP plain radiographs.

Table of Contents:

		Page:
Abstract		
Table of Contents		
Figures and Tables		
Acknowledgements		
Author's Declaration		
Chapter 1	**Introduction to radiographer plain radiograph reporting**	
1.1	**Introduction**	1
1.2	**Role development in Radiography**	4
1.2.1	*What is role development in Radiography?*	4
1.2.2	*How can role development in Radiography be achieved?*	4
1.2.3	*What are current areas of role development in Radiography?*	5
1.2.4	*What is the future of role development in Radiography?*	5
1.3	**Background to radiographer reporting**	6
1.3.1	*The early years*	6
1.3.2	*Recent contentions*	7
1.3.3	*The current position on radiographer reporting*	8
1.3.4	*Background to radiographer reporting*	9
1.4	**Background to image interpretation**	10
1.4.1	*What is a report?*	10
1.4.2	*What is a 'good' report?*	11
1.4.3	*What is observer error and variation?*	12
1.5	**Background for the proposed research**	14
1.6	**Thesis outline**	16
Chapter 2	**Radiographer plain radiograph reporting performance: A systematic review**	
2.1	**Rationale for undertaking the review**	18
2.1.1	*Introduction*	18
2.1.2	*Checking for existing and ongoing reviews*	19
2.1.3	*Advisory Group*	19
2.1.4	*Background research*	20
2.2	**Objectives**	26
2.2.1	*Primary objectives*	26

2.2.2	*Secondary objectives*	26
2.2.3	*Supplementary objectives*	26
2.3	**Searching the literature and retrieving the studies**	27
2.3.1	*Electronic databases*	27
2.3.2	*Handsearching*	30
2.3.3	*Personal Communication*	30
2.3.4	*Reading reference lists*	30
2.3.5	*Grey Literature*	30
2.3.6	*Search limits*	31
2.4	**The selection of studies for inclusion in the review**	31
2.5	**Details of studies included and excluded from the review**	32
2.5.1	*Studies included in the review*	32
2.5.2	*Studies excluded from the review*	34
2.5.3	*Results of study selection process*	35
2.6	**Data extraction form**	36
2.6.1	*Developing the data extraction form*	36
2.6.2	*Piloting the data extraction form*	39
2.6.3	*Applying the data extraction form*	40
2.6.4	*Conclusion*	41
2.7	**Assessing the validity of primary studies**	42
2.8	**Qualitative overview**	44
2.8.1	*Diagnostic accuracy studies*	44
2.8.2	*Diagnostic performance studies*	48
2.8.3	*Diagnostic outcome studies*	51
2.8.4	*Conclusion*	53
2.9	**Quantitative data synthesis (meta-analysis)**	55
2.9.1	*Is meta-analysis feasible?*	55
2.9.2	*Choice of statistical method*	59
2.9.3	*Meta-analyses of radiographer plain radiograph reporting performance*	63
2.9.4	*Conclusion*	72
2.10	**Exploring sources of heterogeneity**	73
2.10.1	*Method of investigating sources of heterogeneity*	73
2.10.2	*Application of Littenberg & Moses extended regression method*	74
2.10.3	*Conclusion*	78
2.11	**Investigating publication bias**	79
2.12	**Interpretation of results**	83

2.12.1	*Assessing the plain radiograph reporting performance of radiographers*	83
2.12.2	*Radiograph reporting performance of selectively trained radiographers compared with other health care professionals*	84
2.12.3	*Radiograph reporting performance of selectively trained radiographers for different categories of patients*	85
2.12.4	*Radiograph reporting performance of selectively trained radiographers for different body areas*	86
2.12.5	*The effectiveness of training programmes for improving radiographer radiograph reporting performance*	88
2.12.6	*Plain radiograph reading performance of radiographers in a red dot or triage role*	89
2.12.7	*Threats to the validity of plain radiograph reading performance studies*	90
2.12.8	*Consideration of factors other than radiograph reading performance*	91
2.13	**Conclusions**	93
Chapter 3	**Development and assessment of methods for measuring radiographer reporting performance**	
3.1	**Introduction**	94
3.2	**Objectives**	94
3.3	**Developing criteria to compare reports for concordance**	95
3.3.1	*Pilot study*	95
3.3.2	*Decision-making criteria used in the primary studies*	98
3.4	**Reliability of the arbiters comparing reports**	102
3.4.1	*Assessment of intra-arbiter reliability*	102
3.4.2	*Assessment of inter-arbiter reliability*	103
3.4.3	*Conclusion about arbiter reliability*	103
3.5	**Acceptability of the reference standard**	104
3.5.1	*Assessment of the reliability of the reference standard*	104
3.5.2	*Assessment of the validity of the reference standard*	105
3.5.3	*Conclusion about the acceptability of the reference standard*	106
3.6	**Conclusion**	106

Chapter 4 **Introducing selectively trained radiographers to an A&E reporting service: A retrospective controlled before and after study**

4.1	**Introduction**	107
4.2	**Objectives**	108
4.2.1	*Primary objectives*	108
4.2.2	*Secondary objectives*	108
4.3	**Setting**	109
4.3.1	*A&E reporting system at site A*	109
4.3.2	*A&E reporting system at site B*	111
4.4	**Methods**	112
4.4.1	*Assessing radiograph reporting performance*	112
4.4.2	*What sample size was necessary?*	116
4.4.3	*Eligibility criteria*	117
4.4.4	*Study sample of reports*	117
4.4.5	*Method of reporting by the reference standard*	118
4.4.6	*Method of comparing reports*	119
4.4.7	*Assessing effect on patient management*	119
4.4.8	*Assessing effect on patient outcome*	122
4.4.9	*Data analyses*	123
4.5	**Results**	124
4.5.1	*Effect on radiograph reporting performance*	124
4.5.2	*Effect on patient management*	126
4.5.3	*Effect on A&E patient outcome*	131
4.6	**Discussion**	133
4.7	**Conclusion**	137

Chapter 5 **Radiographers and Radiologists reporting A&E and GP plain radiographs: A quasi-experimental study**

5.1	**Introduction**	138
5.2	**Objectives**	139
5.2.1	*Primary objective*	139
5.2.2	*Secondary objective*	139
5.3	**Framework for assessing the effects of CSR and radiologist reporting**	140
5.4	**Assessment of diagnostic performance**	143

5.4.1	*Study design*	143
5.4.2	*What sample size was necessary?*	144
5.4.3	*Eligibility criteria*	145
5.4.4	*Selection of radiographs*	146
5.4.5	*Method of reporting by the CSR and radiologists*	148
5.4.6	*Method of reporting by the reference standard*	149
5.4.7	*Method of comparing reports*	149
5.5	**Assessment of diagnostic and therapeutic outcome**	150
5.6	**Assessment of patient outcome**	151
5.7	**Data analyses**	152
5.8	**Results**	153
5.8.1	*Diagnostic Performance*	154
5.8.2	*Diagnostic Outcome*	161
5.8.3	*Therapeutic Outcome*	164
5.8.4	*Patient Outcome*	166
5.9	**Discussion**	168
5.10	**Conclusion**	172
Chapter 6	**A cost analysis of A&E plain radiograph reporting**	
6.1	**Introduction**	173
6.2	**Objectives**	175
6.2.1	*Primary objective*	175
6.2.2	*Secondary objective*	175
6.3	**Methods**	175
6.3.1	*Choice of economic evaluation*	175
6.3.2	*Identification of costs of plain radiograph reporting*	177
6.3.3	*Measurement of costs of plain radiograph reporting*	179
6.3.4	*Valuation of costs of plain radiograph reporting*	179
6.4	**Data collection**	180
6.4.1	*Cost data*	180
6.4.2	*Time taken to report*	180
6.4.3	*Number of radiographs reported*	181
6.5	**Cost analysis of radiographer reporting A&E plain radiographs**	182
6.6	**Results**	183
6.6.1	*The cost of reporting A&E plain radiographs of the appendicular skeleton*	183

6.6.2 *The cost of reporting all A&E plain radiographs* 187
6.7 **Discussion** 188
6.8 **Conclusion** 190

Chapter 7 **Conclusions and future research opportunities**

7.1 **Introduction** 191
7.2 **Discussion about radiographer plain radiograph reporting** 191
7.2.1 *Opportunities and threats to radiographer reporting* 191
7.2.2 *Radiographer plain radiograph reporting* 194
7.2.3 *Recommendations about radiographer plain radiograph reporting* 198
7.3 **Methodological issues about radiographer reporting** 199
7.3.1 *Development of a quality criteria checklist* 199
7.3.2 *Choice of study design* 200
7.3.3 *Assessment of resource management implications* 200
7.4 **Conclusion** 202

Annex 2.1 **Search strategies and references**
Annex 2.2 **Data extraction form**
Annex 2.3 **Checklist results from primary studies**
Annex 2.4 **Results of scoring systems**
Annex 2.5 **Exploring sources of heterogeneity**
Annex 2.6 **Heterogeneity in presence of biases and methodological standards**
Annex 4 **Proformas**
Annex 5 **Proformas and Additional marking criteria**
Annex 6 **Cost Questionnaire and Scenario Analysis**

List of abbreviations

References

Figures and Tables

Chapter 2 Radiographer plain radiograph reporting performance: A systematic review

Figure 2.1	*Classification of studies of plain radiograph reading performance*	23
Table 2.1	*Types of plain radiograph reading performance studies*	23
Figure 2.2	*Framework for review about radiographer radiograph reporting performance*	25
Table 2.2	*Number of studies located from each database*	32
Table 2.3	*Overlap between databases for eligible studies*	33
Table 2.4	*Studies excluded from the review*	34
Figure 2.3	*Study selection process for systematic review*	35
Table 2.5	*Potential biases in radiograph reading performance studies*	37
Table 2.6	*Methodological standards in radiograph reading performance studies*	38
Table 2.7	*Pilot application of data extraction form*	40
Table 2.8	*Reviewer(s) responsible for study appraisal*	40
Table 2.9	*Final application of the quality criteria checklist*	41
Table 2.10	*System for banding studies by quality*	43
Table 2.11	*Diagnostic accuracy studies that assess radiographers only*	46
Table 2.12	*Diagnostic accuracy studies that assess radiographers and radiologists*	47
Table 2.13	*Diagnostic accuracy studies that assess radiographers and A&E staff*	49
Table 2.14	*Diagnostic performance studies that assess radiographers*	50
Table 2.15	*Diagnostic outcome studies that assess radiographers and radiologists*	52
Table 2.16	*Diagnostic outcome studies that assess radiographers and A&E staff*	54
Figure 2.4	*Study selection process for meta-analysis*	57
Table 2.17	*Studies included in the meta-analyses*	58
Figure 2.5	*D-S plots and corresponding ROC curves with DOR*	62
Figure 2.6	*ROC plot of radiographers plain radiograph reporting performance*	63
Figure 2.7	*Sensitivity estimates with 95 per cent CI*	64

Figure 2.8	*Specificity estimates with 95 per cent CI*	64
Figure 2.9	*Symmetric ROC curve of radiographers' radiograph reporting performance*	65
Table 2.18	*DOR based on a single threshold in each primary study*	66
Figure 2.10	*D vs S plot*	66
Figure 2.11	*Summary ROC curves of radiographer radiograph reporting performance*	67
Table 2.19	*Summary sensitivity and specificity for A&E radiographs*	68
Table 2.20	*Summary sensitivity and specificity for radiographs of different body areas*	70
Table 2.21	*Weighted summary ROC models with and without G1*	75
Table 2.22	*Unweighted summary ROC models with and without G1*	75
Table 2.23	*Sensitivity and Specificity by presence of G1*	76
Table 2.24	*Descriptive statistics for F2*	77
Table 2.25	*Pairwise comparisons for F2*	77
Table 2.26	*Weighted model including G1 and F2*	78
Figure 2.12	*Funnel plot for sensitivity*	80
Figure 2.13	*Funnel plot for sensitivity cubed*	80
Figure 2.14	*Funnel plot for specificity*	81
Figure 2.15	*Funnel plot for specificity cubed*	80

Chapter 3 **Development and assessment of methods for measuring radiographer reporting performance**

Table 3.1	*Kappa values*	95
Table 3.2	*Intra-arbiter reliability*	97
Table 3.3	*Definition of reports*	100
Table 3.4	*Intra-arbiter reliability*	102
Table 3.5	*Inter-arbiter reliability*	103
Table 3.6	*Reliability of the reference standard*	104

Chapter 4 **Introducing selectively trained radiographers to an A&E reporting service: A retrospective controlled before and after study**

Figure 4.1	*Site A reporting system*	110
Figure 4.2	*Site B reporting system*	111
Table 4.1	*Sample size calculations*	117
Table 4.2	*Management options*	120
Table 4.3	*A&E plain radiograph reporting performance*	125
Table 4.4	*Test for difference in A&E plain radiograph reporting performance*	125
Table 4.5	*GP plain radiograph reporting performance*	126
Table 4.6	*Test for difference in GP plain radiograph reporting performance*	126
Table 4.7	*Effect on A&E1 patient management*	127
Figure 4.3	*Effect on A&E1 patient management*	127
Table 4.8	*Effect on A&E2 patient management*	128
Figure 4.4	*Effect on A&E2 patient management*	128
Table 4.9	*Effect on A&E patient management*	129
Figure 4.5	*Effect on A&E patient management*	129
Table 4.10	*Effect on GP1 patient management*	130
Table 4.11	*Effect on GP2 patient management*	130
Table 4.12	*Effect on GP patient management*	131
Table 4.13	*Types of patient outcome due to the original report*	131
Table 4.14	*Patient outcome due to the original report*	132
Table 4.15	*Types of patient outcome of consensus review*	132
Table 4.16	*Patient outcome of consensus review*	132

Chapter 5 **Radiographers and Radiologists reporting A&E and GP plain radiographs: A quasi-experimental study**

Figure 5.1	*The evaluative hierarchy as applied to assessing radiographer reporting*	141
Figure 5.2	*Plan of investigation*	142
Table 5.1	*Description of the different sample of radiographs*	144
Table 5.2	*Sample size calculation*	145
Table 5.3	*Body areas within each stratum*	147

Table 5.4	*An example of random permuted blocks within strata*	147
Table 5.5	*Reporting codes*	148
Table 5.6	*A&E management options*	150
Table 5.7	*Potential effect on patient outcome*	151
Figure 5.3	*Flowchart describing the process of evaluation*	153
Table 5.8	*CSR A&E radiograph reporting performance*	154
Table 5.9	*CSR versus radiologist A&E radiograph reporting performance*	155
Table 5.10	*CSR versus radiologist A&E radiograph reporting accuracy*	155
Table 5.11	*CSR and radiologist A&E radiograph reporting accuracy*	155
Table 5.12	*CSR GP radiograph reporting performance*	156
Table 5.13	*CSR versus radiologist GP radiograph reporting performance*	156
Table 5.14	*CSR versus radiologist GP radiograph reporting accuracy*	156
Table 5.15	*CSR and radiologist GP radiograph reporting accuracy*	156
Table 5.16	*CSR and radiologist A&E versus GP radiograph reporting accuracy*	157
Figure 5.4	*ROC curve comparing CSR and radiologists for A&E1*	157
Table 5.17	*CSR and consultant radiologist Az values*	158
Figure 5.5	*ROC curve comparing CSR and radiologists for all experimental groups*	158
Table 5.18	*A&E radiograph reporting performance (excl. FAR)*	159
Table 5.19	*GP radiograph reporting performance (excl. FAR)*	159
Table 5.20	*CSR versus radiologist radiograph reporting accuracy (excl. FAR)*	159
Table 5.21	*A&E radiograph reporting performance (double blind)*	160
Table 5.22	*GP radiograph reporting performance (double blind)*	160
Table 5.23	*CSR versus radiologist radiograph reporting accuracy (double blind)*	160
Table 5.24	*Changes in diagnoses for A&E patients*	161
Table 5.25	*Changes in diagnosis for GP patients*	162
Table 5.26	*Changes in diagnostic confidence for A&E patients*	163
Table 5.27	*Changes in diagnostic confidence for GP patients*	163
Table 5.28	*Changes in therapeutic confidence for A&E patients*	164
Table 5.29	*Changes in therapeutic confidence for GP patients*	165
Table 5.30	*Patient outcome for A&E1*	166
Table 5.31	*Patient outcome for A&E2*	166
Table 5.32	*Patient outcome for A&E in total*	167
Table 5.33	*Patient outcome for GP1*	167

Table 5.34 *Patient Outcome for GP2* 168
Table 5.35 *Patient outcome for GP in total* 168

Chapter 6 **A cost analysis of A&E plain radiograph reporting**

Table 6.1 *Costs per year of plain radiograph reporting* 183
Table 6.2 *CSR and radiologist A&E radiograph reporting times (seconds)* 185
Table 6.3 *Cost of reporting A&E appendicular skeleton radiographs* 186
Table 6.4 *Cost of CSR and radiologist reporting all A&E radiographs* 187

Chapter 7 **Conclusions and future research opportunities**

Table 7.1 *Vacancy rates (%) for medical posts from 1999 to 2003* 193
Table 7.2 *Vacancy rates (%) for allied professional posts from 1999 to 2003* 193

Acknowledgements

This thesis was completed with the support of NHS Northern & Yorkshire Regional Office for which I am grateful.

I am indebted to the staff at the Trust where the primary research was conducted and authors of studies included in the systematic review without whose continuing support the research presented would not have been possible.

I am also indebted to the members of my research advisory group for their time, expertise and encouragement. In particular I am grateful to Christine Godfrey who commented on all the chapters and for David King whose clinical support was invaluable. I also very much thank Seokyung Hahn for helping me to understand the statistical methods and for her comments of selected chapters.

Andrew Scally was also of great help with the design and conduct of the systematic review and was a continuous source of support. I would also like to thank Ian Russell and Daphne Russell for their comments on early draft chapters, Karen Bloor for comments on final draft chapters, and Keith Piper for comments on final draft chapters and for also being a continuous source of support. I am also grateful to Trevor Sheldon who gave me the impetus to complete my thesis and to Anne Burton for proofreading this thesis. I would also like to thank Samantha Stubbs for her loving support and tolerance whilst I completed this thesis.

I would also like to acknowledge the support and help of the clinician's who contributed to the collection of data for the studies included in this thesis:

- Dr Martin Crowe a consultant radiologist who compared reports for concordance in the primary studies.
- Dr David King a consultant radiologist who produced the reference standard report in the primary studies and assisted with the development of the questionnaire used to assess different healthcare professionals' attitudes towards the introduction of radiographer reporting.
- Dr Richard Mannion a consultant radiologist who reported radiographs in the primary studies and judged whether there were changes in the pre- and post-report diagnoses for the different professional groups.
- Dr Nigel Thomas a consultant radiologist who brought the views of a practising clinical radiologist (working in an environment of radiographer plain film reporting) to the systematic review and assisted with data extraction.

- Dr Niall Warnock a consultant radiologist who reported radiographs in the primary studies, helped develop the criteria for comparing reports for concordance and the questionnaire for assessing different healthcare professionals' attitudes towards the introduction of radiographer reporting.
- Mr Ian Crawshaw and Mrs Lorraine Ford, Clinical Specialist Radiographers, who reported the radiographs in the primary studies.
- Dr Anthony Bowker, Dr Julie Cooper, Dr Neil Jenkins, Dr Christine Ivory, Dr Anna Murphy and Dr Mike Porte, consultant radiologists, who reported the radiographs in the primary studies.
- Dr David Longsworth (a GP), Dr Steven Krane (an A&E Specialist Registrar), Dr Peter Williams and Dr Phillipa Rutter (GP registrars). For the primary studies they judged whether differences between reports would have had an effect on clinician's decision making and patient outcome.
- Mr Alistair Jakeman an A&E Nurse who searched A&E patients' case notes and other hospital records. This information was used in one of the primary studies for judging whether differences between reports would have affected patient management and outcome.

I would also like to thank the following people who contributed in various ways to the studies included in the thesis:

- Mr Stephen Boynes a Senior Health Lecturer in the Division of Radiography at the University of Bradford who assisted with the piloting of the data extraction checklist for the systematic review.
- Ms Janette Colclough a Librarian at the University of York who helped to develop the search strategies for the systematic review.
- Mr Darren Earnshaw and Mr Simon Morning, technicians in the Systems and Network Services Department at the Trust involved in the primary research, who provided the data required to randomly select the radiographs to be included in the primary studies.
- Dr Susan Ethell at the Department of Radiography and Imaging Sciences, University of Lancaster who used the AccuROC package to perform the Receiver Operating Characteristic curve analysis.
- Mr Nigel Marsh the Radiology Divisional Manager at the Trust involved in the primary research who provided cost data and Mrs Patricia Thompson the Radiology Administration Manager who liased with various staff on behalf of the author to ensure successful collection and management of data for the primary studies.

Author's Declaration

All the research presented in this thesis was initiated and conducted by the author. In Chapter 2 Mr Andrew Scally (University of Bradford) was responsible for conducting the cluster analysis. In Chapter 5 Dr Susan Ethell (University of Lancaster) used the AccuROC package to perform the Receiver Operating Characteristic curve analysis. Interpretation of the analyses followed discussion with Professor Christine Godfrey and Dr Seokyung Hahn at the Department of Health Sciences. However, the author is completely responsible for the research presented in the thesis.

Stephen Brealey
January 2004

Chapter 1

Introduction to radiographer plain radiograph reporting

1.1 Introduction

In June 1990, the National Health Service (NHS) and Community Care Act was introduced to help address the increasing demand in health care. This Act gave effect to the *Working for Patients* 1989 White Paper, which announced the internal market reforms for the NHS by the Conservative Party government. These reforms resulted in a major re-structuring of the funding and delivery of health care, including a review of consultant contracts and their 'job descriptions', and the formal introduction of medical audit and resource management to improve quality and efficiency. Indeed, the Audit Commission, a body which until then investigated the efficiency of local government, had its brief extended to the NHS. There was also a desire to bring greater satisfaction and rewards to those working in the NHS. A combination of doctor's reflecting on their role and job satisfaction, increased cost-awareness, and the introduction of medical audit, engendered a climate that blurred the distinction of tasks between medical and allied health care, or non-medical, professionals.

Allied health care is delivered by professionals like chiropodists, paramedics, physiotherapists, speech, language and art therapists who support their medical counterparts. In particular, there are diagnostic radiographers who are responsible for the production of high quality images on radiographs or other recording media, using all kinds of radiation. Radiologists are their medical colleagues who are responsible for reading the images produced by radiographers. The service provided by radiographers and radiologists is integral to the front-line diagnosis of patients in the NHS, although while from 1968 to 1991 radiologist's workload increased by 322 per cent, the number of posts increased by only 213 per cent. Subsequently, radiologists in England were reporting only 60 per cent of examinations within two working days [Royal College of Radiologists (RCR), 1993a], and 22 per cent never reported 10 per cent of radiographs (Rose & Gallivan, 1991). The shortage of radiologists meant that the examinations produced by the radiographers were not always reported, and so did not contribute to the diagnosis of patients, an important factor to inform referring clinician's choice of management and, ultimately, patient outcome. Furthermore, an abnormality detected on a radiograph that could have contributed to patient management and outcome but was not reported could have medico-legal implications. Examinations performed but not reported are also a waste of already scarce resources. It is, moreover, unethical and illegal to expose patients to potentially harmful radiation without the benefit of the image being reported.

A potential solution to the problem of increasing radiologist workload and failure to report examinations, was to allow radiographers to report the images that, after all, they were responsible for producing and ensuring were of high quality. Historically, there were many obstacles to radiographers developing a reporting role: radiologist resistance to change; lack of resources to support the necessary education; and lack of training opportunities to allow radiographers to maximise their talents (Paterson, 1995). However, the shortage in radiologists and changes in government policy during the late 1980s and early 1990s helped to remove these barriers and permit radiographers to further develop this role.

In 1996, the College of Radiographers outlined the minimum requirements necessary to implement radiographer reporting [College of Radiographers (CoR), 1996]. This included professionally and academically accredited competence based training, continuing education, and clinical audit to monitor their performance so as to assure quality in clinical practice. The Diploma of the College of Radiographers was also withdrawn in April 1995, with the introduction of an all-graduate entry into the profession. Simultaneous expansion into the higher education sector and the availability of a variety of specialist and advanced postgraduate courses exposed the profession to a research culture in the form of projects generated by under-graduate and post-graduate students. For the profession to sustain its position in diagnostic medical imaging, the Society and CoR (SCoR) also recognised the need to underpin its development with a sound research base. As a result, all radiographers were encouraged to engage in research to be consistent with the very best in medical practice - of which research is clearly seen as a normal and expected part (CoR, 1994). Indeed, the centenary UK Radiology Congress in 1995 saw the launch of the *Radiography* international peer-reviewed journal as a medium for publishing research in Radiography.

In summary, a shortage in medical doctors and the need to address increasing demand in a health care system with scarce resources brought about changes in government policy, promoting a more flexible and creative use of allied health care professional skills. Notably, the change in climate allowed radiographers to develop their reporting role underpinned by clinical audit and the changes in CoR education policy encouraged a research ethos in the profession.

By the mid 1990s, these changes in policy led to research into radiographers developing a reporting role. It was discovered that selectively trained radiographers could report accident and emergency (A&E) plain radiographs to a high level of accuracy (Loughran, 1994a; Robinson, 1996a). However, there were some methodological limitations to these studies and they only assessed the accuracy with which radiographs were reported. In February 1995 two carefully selected radiographers having completed a training programme in plain radiograph reporting began to report A&E plain radiographs of the appendicular skeleton at the Trust where the primary research in this thesis was conducted. This provided an opportunity to undertake further research, underpinned by appropriate methodology, to evaluate the clinical effects of radiographer plain radiograph reporting in terms of accuracy, changes in patient diagnosis, management plans and outcome, and the associated costs. The Trust where this research was conducted is a typical district general hospital which is described as Trust A and consists of two clinical sites (A and B).

This thesis endeavours to answer the question as to whether selectively trained radiographers should report A&E plain radiograph X-ray examinations and the potential for further extending their reporting role.

The aim of the introductory chapter is:

- to discuss role development in Radiography (section 1.2);
- to discuss the background to radiographer reporting (section 1.3);
- to provide background information to image interpretation (section 1.4);
- to provide background information for the proposed research (section 1.5);
- outline of the thesis (section 1.6).

1.2 Role development in Radiography

The purpose of this section is to provide background information to define role development in Radiography, including a definition of role development and how it can be realised in the profession of Radiography and a discussion on how radiographers' roles are currently being developed and what they may be in the future.

1.2.1 What is role development in Radiography?

Role development represents a fundamental change to the professional practice of radiographers and is subject to the provisions of the statutory and professional codes of conduct which govern such practice (CoR, 1996). It is defined as "representing quantitative and qualitative change in the way radiographers contribute to patient management and health care services" and "demands a high level of skill, training, experience and expertise" (CoR, 1997). All role development activities must be underpinned by continuing education and training programmes, incorporating practice and theory related to work-based competencies, and should be accredited both professionally and academically (CoR, 1996). Radiographers already play an integral role in front-line diagnosis. Such activities should increase their job satisfaction and further develop their professional standing (Hughes et al, 1996).

1.2.2 How can role development in Radiography be achieved?

The developing role of radiographers can be achieved through 'skill-mix' initiatives. A skill is defined as "an act or series of acts in which instruction and practice are required to achieve a level of competence and should be exercised effectively and efficiently without supervision" (RCR, 1993a). SCoR are driving current initiatives to train qualified radiographers at postgraduate level and to include training in degree syllabuses in aspects of what was formerly considered the province of radiologists.

1.2.3 What are current areas of role development in Radiography?

The RCR (1993a) proposed ways in which the role of the radiographer may be developed, including ultrasound scanning (e.g. Doppler techniques and the recording of technical observations), some contrast media examinations (e.g. barium examinations) and intravenous injections (e.g. radiopharmaceuticals and contrast media). There is now evidence that radiographers can provide preliminary reports of abdominal ultrasound examinations (Bates et al, 1994) and successfully perform barium enemas (Mannion et al, 1995). They can also undertake the injection of radiopharmaceuticals in nuclear medicine (NM) and of contrast media in computed tomography (CT), magnetic resonance imaging (MRI) and urography (Robinson, 1996). As discussed in more detail in section 1.3, radiographers now provide both verbal and written reports for A&E radiographs and for other imaging examinations such as ultrasound, MRI, CT, NM (CoR, 1997).

1.2.4 What is the future of role development in Radiography?

It took time to overcome the status quo, set early in the 20th century, before radiologists were willing to delegate some of their more routine tasks to radiographers in the interest of utilising their specialist skills more effectively (Craven & Barber, 1995). Indeed, the future role of radiographers partly depends upon radiologists developing the use of interventional techniques, high technology, and sub-specialisation for which they will require more time free from their current routine work.

As radiologists are beginning to delegate some of their traditional tasks radiographers are doing likewise, and helpers or aides are beginning to undertake many of the tasks they previously performed. It may be that, similar to radiologists leaving behind more routine elements of their profession without losing control over them, radiographers will no longer be able to retain the more routine aspects of theirs (Craven & Barber, 1995). There is also a wide range of occupational groups, including midwives, cardiac technicians and vascular technologists that perform ultrasound imaging examinations (Paterson & Price, 1996) and evidence of emergency nurse practitioners' reading A&E radiographs (Meek et al, 1998; Remedios et al, 1998). These initiatives in other professions all impinge on radiographers' practice.

The continuing need for growth in skills in Radiography is imperative to the retention of their professional identity and status in a dynamic health care climate (O'Connor, 1996). Further developing the role of radiographers will allow Radiography to consolidate its professional status and address future technological and employment challenges. Indeed, the recent government initiatives described in the *NHS Plan* [Department of Health (DH), 2000] and elaborated on in *Meeting the Challenge* (DH, 2000) aim to ensure that the continuing development of the allied health professional role is successful and enduring. The government aims to invest in and reward staff by modernising pay structures, increasing earnings and improving staff working lives in the form of advanced practitioner and consultant posts. The future challenge to radiographers is to fit the criteria for these posts by demonstrating expertise and leadership in the research, clinical and professional environment.

1.3 Background to radiographer reporting

This section provides a summary of the historical background to radiographer reporting.

1.3.1 The early years

The subject of non-medically qualified staff reporting radiographs has been debated and contested almost since the discovery of X-rays by Röntgen in 1895. In the early years of radiation medicine, there was no clear distinction between Radiography and Radiology and the terms were used interchangeably. Radiographers would routinely comment on X-ray plates without medical supervision (Larkin, 1983), however, this was frowned upon by some members of the medical profession (Arthur & Muir, 1909) and within ten years of Röntgen's discovery, moves were made to establish boundaries between medical and non-medical practitioners (CoR, 1995). The debate continued until 1925, when the Society of Radiographers adopted a special resolution restricting its members from giving any form of report on an examination, although in certain circumstances radiographers could describe the appearances seen "to such an extent as may be necessary to assist in making a diagnosis" (Moodie, 1970). Subsequently, Radiography was reduced to mean the production of radiographs and was practised by non-medically qualified or technical personnel: radiographers working at the behest of medical practitioners. Radiology came to mean the medical interpretation of radiographs and became the exclusive domain of medically qualified staff (Paterson & Price, 1996).

1.3.2 *Recent contentions*

Although in many situations it is recognised practice that radiographers comment upon images they produce, it was not until the 1970s before any formal change relative to reporting was proposed. Swinburne (1971) was probably the first to suggest an investigation into 'pattern recognition', whereby a radiographer could identify whether a radiograph is 'normal' or 'abnormal', without prolonged, complex training. Berman et al (1985a) were the first to perform a prospective evaluation of a scheme whereby radiographers marked A&E radiographs with a red dot to alert casualty officers to the possible presence of abnormalities. Radiographers and casualty officers were found to have equivalent error rates when reading A&E radiographs. The 'red dot system' or similar flagging systems are now regarded as very useful in some departments. Two recent national surveys have identified that such a system is employed in 150 and 162 hospitals respectively (Paterson, 1995; Price et al, 1999).

However, there has been considerable concern about developing the red dot system into a written, radiographic report. Fielding (1990) agreed with red dot systems and acknowledged that the reporting of A&E radiographs makes a significant contribution to the workload of many radiologists, but he also believed that the contentious issue of making reports must remain the province of the radiologist. Renwick et al (1991) ruled out reporting A&E radiographs by radiographers, but did recognise that the evidence from their research was based on unselected radiographers of varying levels of expertise, none of whom had been formally trained in fracture recognition. Not until Saxton's controversial editorial in 1992, was it suggested that with training, suitable radiographers could undertake reporting in such areas as fracture recognition on A&E radiographs. Subsequently, moves were made to evaluate the feasibility of radiographers contributing to plain radiograph reporting services with three separate, unrelated, but almost simultaneous initiatives in Leeds, Macclesfield, and Canterbury. All three studies concluded that experienced radiographers who receive supplementary training may be introduced into a reporting rota for A&E skeletal examinations. These initiatives were followed quite rapidly by the development of reporting training programmes for radiographers and, by December 1996, five Higher Education Institutions were offering postgraduate programmes: Bradford, Hertfordshire, Leeds, South Bank, Canterbury Christ Church College and Salford. To improve the awareness of what was happening in local situations, the national Special Interest Group in Radiographic Reporting (SIGRR) was established in 1996, providing a forum for parties with an interest in radiographer reporting (Cunningham, 1997).

Since then, there has been an increase in the number of education programmes available to prepare radiographers for a reporting role (Prime et al, 1999). Moreover, a comparison between two national surveys clearly indicates an increase in the number of radiographers reporting A&E radiographs. Paterson (1995) identified radiographer reporting at four Trusts only, whereas Price et al (1999) found that 37 Trusts stated radiographers were involved in reporting. This is evidence that local initiatives to train radiographers to report are affecting practice on a national scale. The danger is that in the absence of robust evidence this might become a natural duty of radiographers rather than a role extension, further supporting the need for rigorous evaluation of radiographer plain radiograph reporting to justify their already developing role in this area.

1.3.3 The current position on radiographer reporting

During the 1990s, the CoR, RCR and Audit Commission issued papers reflecting and encouraging relaxation of restrictions on radiographer reporting.

Understandably, the CoR were always enthusiastic about radiographers providing a report on image appearances, beginning with a statement to this effect in 1993. These sentiments were re-iterated the following year in the *Code of Professional Conduct*, which stated "radiographers may provide a verbal comment on image appearances and should provide a written report to the clinician" (CoR, 1994). In 1995, a paper discussed the importance of training in the reporting of radiographs (CoR, 1995) and in *Reporting by radiographers: A vision paper*, the belief was again expressed that all radiological examinations carried out by radiographers, irrespective of the imaging modality used, should receive a radiographer report (CoR, 1997).

In contrast, the RCR began with the statement in 1993 that "it would not be appropriate to expect a non-medical practitioner to offer a medical opinion on a radiograph or procedure" (RCR, 1993b). However, by 1995 they recognised that there were "insufficient fully trained radiologists to undertake all the procedures and report all the examinations" and explored "alternative ways of providing reports, principally in relation to plain radiographs" and the potential of delegation to non-medical staff, such as radiographers (RCR, 1995a). In 1998, a joint statement by the RCR & CoR described how the reporting of radiological images could be properly delegated to non-medical staff.

More recently, the SIGRR have published a document that builds on the CoR (1997) *Reporting by radiographers: A vision paper.* The SIGRR (2002) paper provides further guidance on policy and practice for staff involved in this skill mix, to reflect the increase in number of radiographers involved in reporting and the number of education programmes available.

1.3.4 Background to radiographer reporting

At a typical general hospital the A&E department is responsible for 27 per cent of the radiographic examinations and general practitioner (GP) examinations account for 21 per cent (Audit Commission, 1995). At the general hospital where the primary research for this thesis was to be conducted, which shall be called Trust A, the corresponding workload in 1997 to 1998 amounted to 23 per cent A&E examinations and 19 per cent GP examinations.

Radiographer reporting was introduced at Trust A in February 1995, as a result of a project funded between 1992 and 1995 by the NHS Executive and the then Yorkshire Regional Health Authority, at a cost of £85,000. Two radiographers at Trust A and two from another local hospital were selected, based on their experience and competence, to undertake a training programme in plain radiograph reporting. The aim was to enhance the radiographers' skills in the reporting of skeletal, chest and abdominal radiographs, to ensure that all radiographs could continue to be reported, and to contain costs. The results of the study demonstrated that the trained radiographers performed better than comparison groups of untrained radiographers and trainee radiologists when reporting plain radiographs for all areas of the body (Personal Communication from Jean Wilson, 1999). The radiographers at these Trusts were introduced to their respective A&E reporting services, with the caveat that their performance was regularly monitored to ensure a consistent level of quality. Internal agreements were made between the radiologists and A&E consultants at Trust A, allowing the radiographers to report under the new title of Clinical Specialist Radiographers (CSRs).

At present, the CSRs report only a selected sample of radiographs, as do selectively trained radiographers at other district hospitals. These are A&E plain radiographs of the appendicular skeleton that have been read by medical staff in Casualty and then returned to the X-ray Department for review. They exclude the radiographs of patients with overt fractures, who are referred to the next available fracture clinic and subsequent follow up radiographs. Approximately 90 per cent of the radiographs reported have negative or equivocal findings and the remaining 10 per cent are subtle, positive findings. At Trust A, as at other hospitals, there was the potential to extend radiographer reporting to include axial, as well as chest and abdominal A&E radiographs, or even to other categories of patients, but there is uncertainty as to whether such programmes should be extended.

1.4 Background to image interpretation

An understanding of the complexities involved in defining a report and illustrating the problems with measuring reporting performance is important to the appreciation of current controversies concerning radiographer reporting. This section will define a report, outline what constitutes a 'good' report, and discuss the salient issues regarding observer error and variation when interpreting images.

1.4.1 What is a report?

The difference between 'pattern recognition' used in red dot systems and 'reporting' is that the latter involves the translation of the observed abnormality into an explanation of the findings in terms of pathology. In many cases, further commentary on the significance of the results in the context of the individual patient is necessary (Robinson, 1996b).

Pattern recognition technique describes a process that requires no systematic visual analysis or disciplined effort. In contrast, reporting involves an analytical approach to the perception of image features, followed by synthetic processes of deduction or induction to achieve an understanding of the pathological basis of the abnormalities shown, and their medical significance (Robinson, 1998). Furthermore, the reporting process has two elements: the descriptive report, which involves the interpretation of the radiograph appearances; and the medical report that includes an opinion on the further medical management of a patient (RCR & CoR, 1998). Hence, a report is an "expert" opinion expressed as a verbal or written description and interpretation of image appearances based on past experience and current observation.

The descriptive and interpretational elements of a report can be combined under the general term 'findings', involving the three steps of perception, interpretation and diagnosis. Perception requires visual recognition of the image features and mental comparison with historically recollected data describing normal and abnormal appearances. Interpretation calls for an understanding of the mechanisms of disease or trauma which cause abnormal appearances, and an understanding of the range of normal variation and changes with age. To diagnose, the observer must deduce from the radiograph appearances, the nature, extent and underlying disease process. A further element of some reports is the recommendation of future steps in patient management, such as suggestions for additional imaging procedures or proposals for interventional therapy (Robinson, 1996b). An advantage of the A&E environment is the relatively limited range of pathology, particularly in musculo-skeletal areas of the body, compared with the much wider range of possible findings in unselected inpatient or outpatient examinations (Robinson, 1998).

1.4.2 What is a 'good' report?

The reporting of images is an exercise in communication. In order to succeed, the right message must be sent at the right time to the right person (Robinson, 1996b). At present, the indicators used for measuring the quality of a report, which is synonymous with the quality of a reporting service, relate to availability and content. The former emphasises the speed of report production from the time the examination was performed to its receipt by the referring clinician, and to a lesser extent the proportion of images reported (RCR, 1995a; ACR, 1995). The content of a report focuses on intrinsic dimensions of report quality in terms of reliability, accuracy and readability (RCR, 1995a; Sierra et al, 1992). The other principal attributes are clarity, brevity, clinical correlation and the appropriate reporting of pertinent negative findings (Lafortune et al, 1988).

To satisfactorily verify the quality of a report, which is also an assessment of an individual's performance, it must be compared with a standard, making its validity dependent on the veracity of the reference standard (Robinson, 1997). But the process of deriving an incontrovertible standard to help assess the accuracy of a report poses several difficult methodological problems. For instance, a report is described as an opinion only, so by definition, it admits that doubt exists, since when certainty is established opinion becomes superfluous (Robinson, 1997). This problem is illustrated by the use of qualifiers to convey observer's uncertainty in recording the absence or presence of an abnormality, the severity or degree of abnormality, and progression of disease with time in relation to treatment (Robinson & Fletcher, 1994). Quality, is also, by definition, unmeasurable, since it represents precisely those attributes of an entity which are indefinable in quantitative terms. No robust methodology has yet been developed for the assessment of how well cognitive tasks, such as interpretation of radiographs, are performed (Robinson, 1997).

These problems are not insurmountable. Reports can be graded to reflect the various qualifiers and analysed accordingly, and the individual intrinsic dimensions of reports can be appraised, which in totality reflect the quality of a report. In the absence of explicitly defined standards, it has been suggested that a useful guideline for clinical practice may be the medico-legal benchmark: an acceptable report is indistinguishable from that made by an "average" practitioner (Robinson, 1999). Using this approach, concordance between reports can be assessed by measuring the level of agreement or reliability between individuals or 'observers' of equal competence. Alternatively, the reporting accuracy of an observer under evaluation could be measured in comparison with a reference standard report. This is generated by agreement between a panel of independent consultant radiologists or the report of a single, experienced consultant radiologist validated by appropriate clinical follow-up. Since some observer variation is idiosyncratic or due to random mistakes, independent agreement between a group of observers should provide a better standard than a single expert. However, reproducible results are not necessarily accurate, as all observers could agree on a finding and all of them could be wrong (Robinson, 1997).

1.4.3 What is observer error and variation?

Interpretations that differ from the view of an independent panel of "experts" are regarded as errors; where experts fail to achieve agreement, differing reports is "observer variation". An error reflects an inaccurate interpretation in comparison with the standard opinion of expert radiologists, where the validity of the interpretation is dependent upon the veracity of the standard. When experts fail to agree, and there is no standard by which to measure their performance, this is considered a source of variation (Robinson, 1997).

The concept of error and intra- and inter-observer variation, or variation within and between observers respectively, is not new. Over 50 years ago, it was recognised that the "human equation" resulted even in experts exhibiting enormous variations in their ability to be consistent with themselves and others equally competent. This element of uncertainty extends to all branches of medicine (Garland, 1949). Cognitive psychology recognises that human beings behave differently and are by no means neutral or passive toward incoming information (Neisser, 1967). Similarly, clinical judgement in the context of image interpretation is not objective and passive but a subjective and active process of synthesising information, susceptible to imprecise or inconsistent deductions reflecting human heterogeneity.

It is believed that errors and variation in interpretation now represent the weakest aspect of clinical imaging (Robinson, 1997), an opinion substantiated by the discovery that 70 per cent of legal cases arising within Radiology departments are a consequence of alleged diagnostic errors (Berlin, 1995). These errors can arise from poor technique, failures of perception, lack of knowledge and misjudgement (Berlin, 1996a, b, c, d). The majority of litigation cases arise from the failure to diagnose breast cancer on mammography, lung cancer on chest radiographs, and fractures on skeletal radiographs [Physician Insurers Association of America (PIAA), 1997]. Several systems of classifying errors have been developed and Smith was probably the first to develop such a scheme in 1967. This was later updated by Renfrew et al (1992) who classified the causes of error as limitations of technique, misleading or incomplete clinical data, unavailability of previous studies or reports, false positive errors (over-calls), misinterpretation of perceived findings, and misses due to the phenomenon of "satisfaction for search" in which subtle findings are more likely to be overlooked if overt abnormalities are also present. Kundel (1989) also distinguished between perceptual and cognitive errors. The former occur when image features, though recorded, are not appreciated - the failure of an observer to correctly describe the image appearances. A cognitive, or reasoning error, occurs when image features, though appreciated, lead to erroneous conclusions, so having identified an abnormality the observer incorrectly interprets what the abnormality is. This eclectic array of potential sources of error exemplifies the problems associated with image interpretation.

Observer variation in plain radiograph reporting is also substantial. A recent study examined the variation between three experienced observers reporting the three major types of plain radiograph examination: skeletal, chest and abdomen. Concordance between all three readers was found in 74 per cent, 61 per cent and 51 per cent of radiographs respectively (Robinson et al, 1999). Observer variation should also be considered when different diagnostic methods are compared; in many cases, the difference between observers outweighs the difference between techniques (Kido et al, 1993).

Finally, the magnitude of observer variation must be considered when designing assessment techniques and setting quality standards for the reporting of radiographs (Robinson et al, 1999). This problem is exacerbated by variation in judgements about whether reports are concordant. Development of objective and reliable criteria to minimise variation in the measurement of performance is important so that when an intervention for improving performance (e.g. a training programme) is evaluated, change can be attributed to the policy rather than reflecting inconsistencies in measurement. Assessment of observer performance should be underpinned by scientific principles if unbiased, valid and reliable results are to be collected.

1.5 Background for the proposed research

Both the RCR (1995a) and the American College of Radiology (ACR, 1995) state that all radiographic examinations should be accompanied by a timely, accurate and appropriate written report. Frequent emphasis has been placed on the clinical contribution of the radiologist's report in the management of A&E patients (de Lacey et al, 1980; Thomas et al, 1992; Wardrope & Chennels, 1985) and the Audit Commission (1995) also supports this view. They recommend that Radiology departments institute 'hot' reporting systems, allowing reports on basic examinations to be available before the patient leaves the department, as a delay in the reporting of radiographs may diminish the effectiveness of patient management.

As discussed earlier, an increase in radiologist workload from 1968 to 1991 has meant that not all radiographs are reported. The Audit Commission's survey in 1995 drew attention to the degree of clinicians' dissatisfaction, including those from A&E, with some aspects of the reporting service provided by Radiology departments. In particular, it noted that reports were not provided for all examinations and that a significant percentage was not received in time to influence patient management. This raised potential medico-legal issues, particularly important within the field of A&E medicine, as two-thirds of all claims concern radiographs and over half relate to missed abnormalities or difficult interpretations (Capsticks Solicitors, 1994).

In view of the difficulties in fulfilling recommended standards, attention has focused on radiographers reporting under supervision on selected groups of examinations (Robinson et al, 1999). When considering the transfer of responsibility for reporting selected cases from radiologists to radiographers, it is clearly essential to ensure service quality is not adversely affected. Studies have demonstrated that selectively trained radiographers can achieve the same standard of accuracy as radiologists when reporting A&E radiographs (Loughran, 1994; Robinson, 1996a). Radiographer reporting has the potential to alleviate radiologists' reporting workload, allowing them more effective use of their time by performing other more specialist and complex investigations. It may also increase the potential for reporting a higher proportion of radiographs and in a more timely fashion. This could have a beneficial effect on clinician and patient satisfaction, patient management and outcome, securing improved service quality (Audit Commission, 1995). The job satisfaction and skill of radiographers would be enhanced and the professional profile of Radiography further consolidated. Finally, managers view skill mix or 'workforce re-profiling' as a way to make substantial savings on unit labour costs (Kletzenbauer, 1996). It is believed that the wider deployment of radiographers in a developed reporting role will bring enormous benefits to the patient, and has the capacity to revolutionise the cost-effective management of the patient in clinical radiology and other imaging dependent services (CoR, 1997).

Most of the above is conjecture. As yet, the effects of introducing radiographers to multiple facets of the reporting service such as radiologists reporting workload, the proportion of radiographs reported, the timeliness of reports and acceptability to health care professionals and patients awaits rigorous evaluation. Those studies that have assessed radiographer radiograph reading performance are susceptible to biases that could overestimate their accuracy. No economic evaluation has been conducted. Nor have the chain of events that follow report availability and content, or report quality, been assessed. For example, timely, accurate and coherent reports are necessary for reassuring referring clinicians by improving their confidence in their diagnosis and contributing to the decision to undertake another diagnostic test, which may have economic implications and determine whether a patient is further exposed to radiation. The report will also influence the choice of patient management, which may ultimately affect patient outcome.

In summary, a comprehensive assessment of the cost and benefits of using selectively trained radiographers needs rigorous examination of as many of these dimensions as is feasible. This can be achieved with a variety of research methodologies.

1.6 Thesis outline

The aim of the thesis was to determine whether selectively trained radiographers should report plain radiograph X-ray examinations. First, it was appropriate to conduct a systematic review to synthesise the existing evidence about radiographer radiograph reporting. Before conducting the primary research, it was necessary to develop and assess the criteria and standards used to measure reporting performance to help ensure collection of reliable and valid data. Primary studies were then designed to evaluate the clinical effects of radiographer radiograph reporting and associated costs. Finally, it was important to discuss the evidence from these studies to inform policy, influence good practice and direct research. The following is a summary of the objectives and contents of each chapter.

Chapter 2

A systematic review of radiographer and other health care professionals plain radiograph reporting performance for different body areas and patient types.

Data was also collected in detail on threats to study validity and whether there was evidence about the clinical effects of radiographer reading performance on, for example, patient diagnosis and choice of patient management, and the associated costs. This was to help inform the design of the primary studies in Chapters 4 to 6.

Chapter 3

A feasibility study to develop the decision-making criteria used to compare reports for concordance in the primary studies presented in Chapters 4 and 5.

This was followed by an assessment of the consistent application of these criteria as well as the acceptability of the reference standard. The development of these methods for measuring reporting performance was to help ensure that valid and reliable data was collected from the primary studies.

Chapter 4

A controlled before and after study to assess the effect of introducing selectively trained radiographers to an A&E reporting service.

This included an assessment of radiograph reporting accuracy and effect on patient management and outcome: a pragmatic study reflecting normal clinical practice.

Chapter 5

A quasi-experimental study to assess the potential for extending radiographer's reporting role.

The study involved assessing the performance of the selectively trained radiographers and a group of consultant radiologists at Trust A, in comparison with a reference standard when reporting A&E and GP plain radiographs for all body areas. An explanatory study to assess the efficacy of the two professional groups ability to independently report in a controlled environment, it included measuring the effect of reports on the diagnosis, management and outcome of the patient.

Chapter 6

An analysis of the cost of introducing radiographer reporting plain A&E radiographs of the appendicular skeleton and the cost of extending their reporting role to include the remaining body areas.

The analysis was supplemented by an in-house survey that qualitatively explored the consequences of introducing radiographer reporting at Trust A on different professional's workload.

Chapter 7

The aim of Chapter 7 was to use the evidence presented in this thesis to discuss the conclusions that can be drawn about selectively trained radiographers reporting plain radiographs and what future research is necessary.

All primary research received ethical approval from the Local Research Ethics committee. Data collection adhered to the Data Protection Act, 1984.

Chapter 2

Radiographer plain radiograph reporting performance: A systematic review

2.1 Rationale for undertaking the review

2.1.1 Introduction

Historical and current contentions relevant to non-medical staff reporting radiographs were discussed in Chapter 1. It was therefore timely to conduct a systematic review with the *primary objective* of assessing radiographer plain radiograph reporting performance to help establish whether the increasing trend in this skill mix is justified (Paterson, 1995; Price et al, 1999). Other *secondary* objectives worthy of consideration were: accuracy of selectively trained radiographers reporting compared with other health care professionals; accuracy of reporting different categories of patients and body areas; and effectiveness of training programmes for improving radiographer reporting performance.

There were also *supplementary* issues that would be useful to address when synthesising the evidence about radiographer plain radiograph reporting performance. These included: assessing radiographers reading plain radiographs in a red-dot or triage role; identifying threats to study validity; assessing the clinical effects of radiographer reading performance on patient diagnosis, management and outcome, and the associated costs.

Although red-dotting or triaging radiographs only involves the use of pattern recognition techniques that require limited understanding of the pathological basis of abnormalities shown on radiographs, this nevertheless contributes to clinician's decision-making and subsequently patient outcome and costs. A recent survey showed that 162 Trusts use radiographers in this role (Price et al, 1999). It was therefore appropriate to also synthesise the evidence about radiographer's performance in this role.

The complexities of reading images and the associated variation between observers were discussed in Chapter 1. The suggestion was that assessing radiographer's reporting performance could be extremely subjective and prone to bias, so identification of threats to study validity should be considered when appraising such studies. This should help inform the synthesis of the results of studies included in this review and the design and conduct of the primary studies presented in Chapters 4 to 6.

Again, as discussed in Chapter 1, there seemed to be an absence of evidence about the clinical effects and costs of radiographer reporting. It was therefore also appropriate to collect evidence about this as a supplementary objective of the review. This in turn could help to justify the design of the primary studies in this thesis.

2.1.2 Checking for existing and ongoing reviews

Given that the rationale exists for undertaking a systematic review, the following databases were searched and the results provided in Annex 2.1:

- MEDLINE; and
- NHS Centre for Reviews and Dissemination (CRD) at the University of York, Database of Abstracts of Reviews of Effectiveness (DARE).

These searches did not identify a review. The Cochrane Collaboration and the Special Interest Group in Radiographer Reporting (SIGRR) informed the author that to their knowledge a review did not exist, neither was one in preparation or commissioned.

2.1.3 Advisory Group

The advisory group comprised the author and members of Department of Health Sciences with extensive experience in conducting reviews. Two reviewers were recruited (one from Radiography and the other Radiology) to provide subject area advice, help identify unpublished data, assist with data extraction and reflect the major professional perspectives. Both reviewers are members of the SIGRR and have research experience in the subject area. The radiographer is a Lecturer at the School of Health Studies, University of Bradford and the consultant radiologist is based at North Manchester General Hospital and has a special interest in radiographer role development and education. They are also potential users of the review, so can facilitate its dissemination and implementation. Finally, a Senior Health Lecturer at the School of Health Studies, University of Bradford, helped to pilot the data extraction form.

2.1.4 Background research

Results of the background research to conducting the review are presented here, including discussion of the scope of the review, formulation of its objectives, the type of study designs used to assess radiographer plain radiograph reporting performance and the framework for conducting this review.

2.1.4.1 Scope of the review

When performing the background research, a broader search identified studies that assessed radiographers reading images other than plain radiographs. For example, when reporting routine Computed Tomography (CT) head scans compared with a consultant radiologist an experienced radiographer performed as well as a group of five senior registrar radiologists (Craven & Blanshard, 1997). Studies also demonstrated that selectively trained radiographers perform at a similar level to radiologists when reading mammograms (Pauli et al, 1996; Haiart & Henderson, 1991) and read abdominal ultrasound examinations at a high level of accuracy (Bates et al, 1994). Evidence also emerged of other health care professionals, such as casualty officers (Berman et al, 1985), radiologists (de Lacey et al, 1980) and nurse practitioners (Overton-Brown & Anthony, 1998; Meek et al, 1998), reading radiographs.

Although this provided the possibility of broadening the scope of the review to include all health care professionals reading a variety of images, it was decided to retain the focus on radiographer plain radiograph reporting performance, since this was the aim of the thesis. Expanding the scope of the review would also diminish the rigor with which it could be conducted in terms of, for example, identifying all relevant studies and extraction of data by two independent reviewers.

2.1.4.2 Formulating the objectives of the review

In the interests of formulating objectives and identifying possible effect modifiers, some relevant published studies were located. They were of varying quality, according to traditional hierarchies of evidence (Deeks et al, 2001) and methodological papers on evaluating diagnostic tests (Jaeschke et al, 1994; Reid et al 1995) and were undertaken in different settings, for different patient types and body areas. They also assessed radiographers reading radiographs in different roles. Issues concerning the effect of a report on a clinician's diagnosis, therapy decisions, patient outcome and cost-effectiveness were consistently ignored, confirming the need to meet the objectives described in the introduction to the rationale for the review.

2.1.4.3 Type of study designs

It is important when conducting a review to consider the type of study designs used to address its objectives, as this can influence decisions on the validity of the evidence and the conclusions that can be drawn. Background research identified that studies of radiograph reporting performance involve observers (e.g. radiographers) reading a sample of radiographs under exam conditions or during clinical practice. An arbiter (i.e. health care professional) then judges whether the reports made by the observers are concordant with a reference standard (e.g. consultant radiologist), with resulting data used to calculate statistics like sensitivity and specificity.

Studies conducted in different settings were very different in design. Those performed under exam conditions were more explanatory in design and assessed the efficacy with which radiographers read plain radiographs. A mix of normal and abnormal radiographs were carefully selected with the abnormalities covering a range of pathology, body areas and degrees of conspicuity. A robust reference standard such as a double/triple blind consultant radiologist report was often developed, against which to compare radiographer reports, thus ensuring the radiographers read radiographs to a high level of accuracy before their introduction to clinical practice. Subsequently, samples of fewer than one hundred radiographs were often used.

In contrast, studies conducted during clinical practice were more pragmatic in design and assessed the effectiveness with which radiographers read a series of radiographs ranging from several hundred to several thousand, often compared against a single consultant radiologist as the reference standard.

For the purpose of the review, it was decided to call the studies performed under exam conditions 'Diagnostic Accuracy' studies and those conducted during clinical practice 'Diagnostic Performance' studies. The titles were used because the aim of the studies performed under exam conditions was to use a robust reference standard to ensure that radiographers reported *accurately* in controlled conditions before being introduced to clinical practice. Those studies conducted during clinical practice were more concerned with assessing radiographer *performance* when reading a larger sample of radiographs, for which it was not feasible to generate a robust reference standard.

In both settings, some studies assessed not only *radiographer* reading performance but also the performance of other professional groups. To assist data synthesis, it was beneficial to categorise those studies conducted during clinical practice by whether or not they assessed only *radiographer* performance. This was to distinguish between studies that presented findings about radiographer performance from those that were assessing how radiographers could substitute or complement another professional group. 'Diagnostic Performance' studies which assessed both radiographer and other health care professionals performance were labelled 'Diagnostic Outcome' studies, under the assumption that if one professional group read radiographs more accurately than the other, it would lead to a change in clinician behaviour that could affect patient *outcome*. The procedure was not applied to studies conducted during exam conditions, as they only presented the accuracy with which different professional groups read plain radiographs under controlled conditions, and were not designed to be generalised to clinical practice.

Figure 2.1 summarises how the studies of radiographer plain radiographer reading performance were classified and Table 2.1 defines the three different types of study design.

Figure 2.1 Classification of studies of plain radiograph reading performance

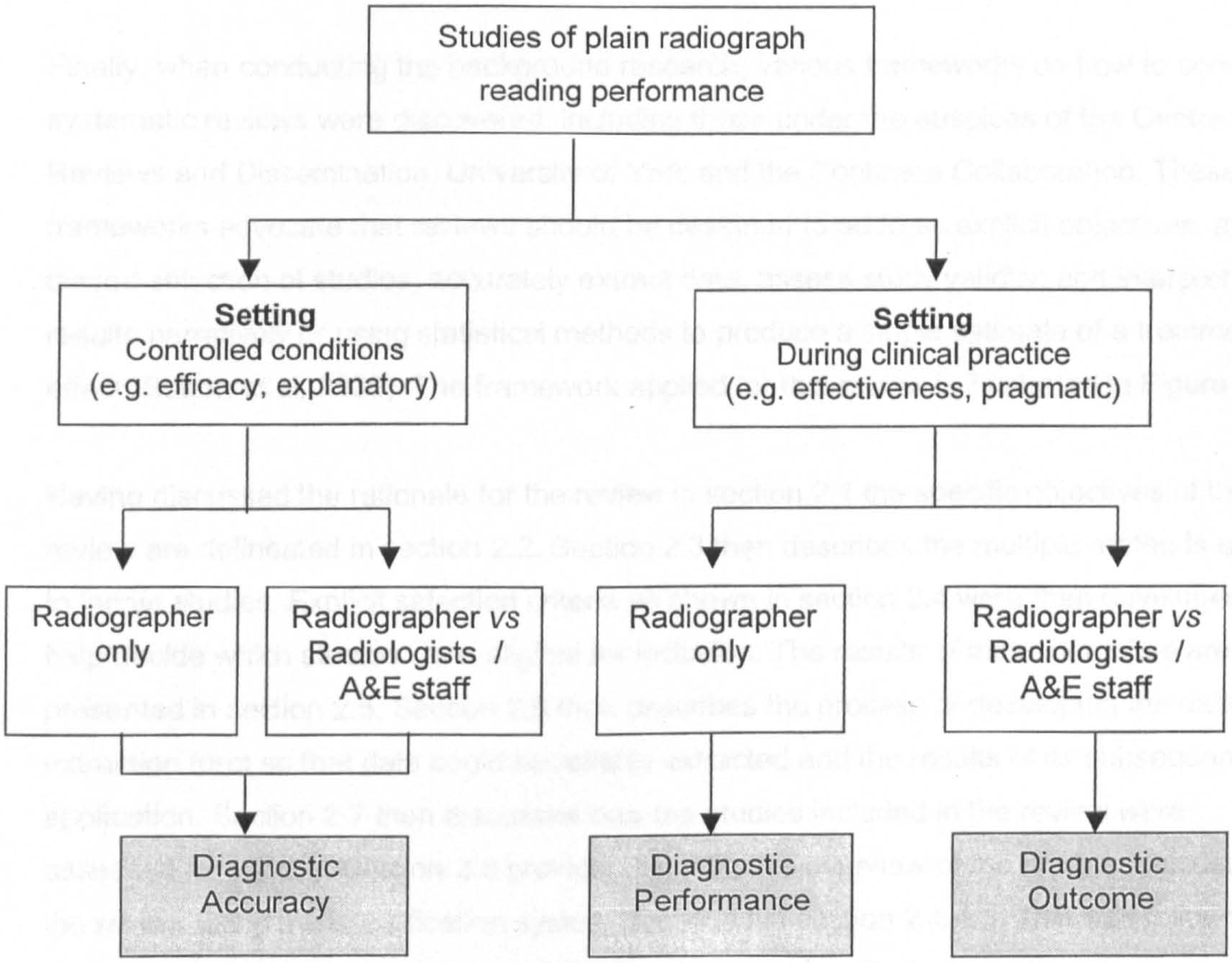


Table 2.1 Types of plain radiograph reading performance studies

Type	**Description**	**Example**
Diagnostic Accuracy	To assess the radiograph reading performance of one (or more) group of observers in controlled conditions	Radiographers reporting on a validated bank of radiographs on a postgraduate course
Diagnostic Performance	To assess the radiograph reading performance of one group of observers during clinical practice	An audit of radiographers radiograph reading performance
Diagnostic Outcome	To assess the radiograph reading performance of two (or more) groups of observers during clinical practice	A comparison of radiographers and radiologists radiograph reading performance

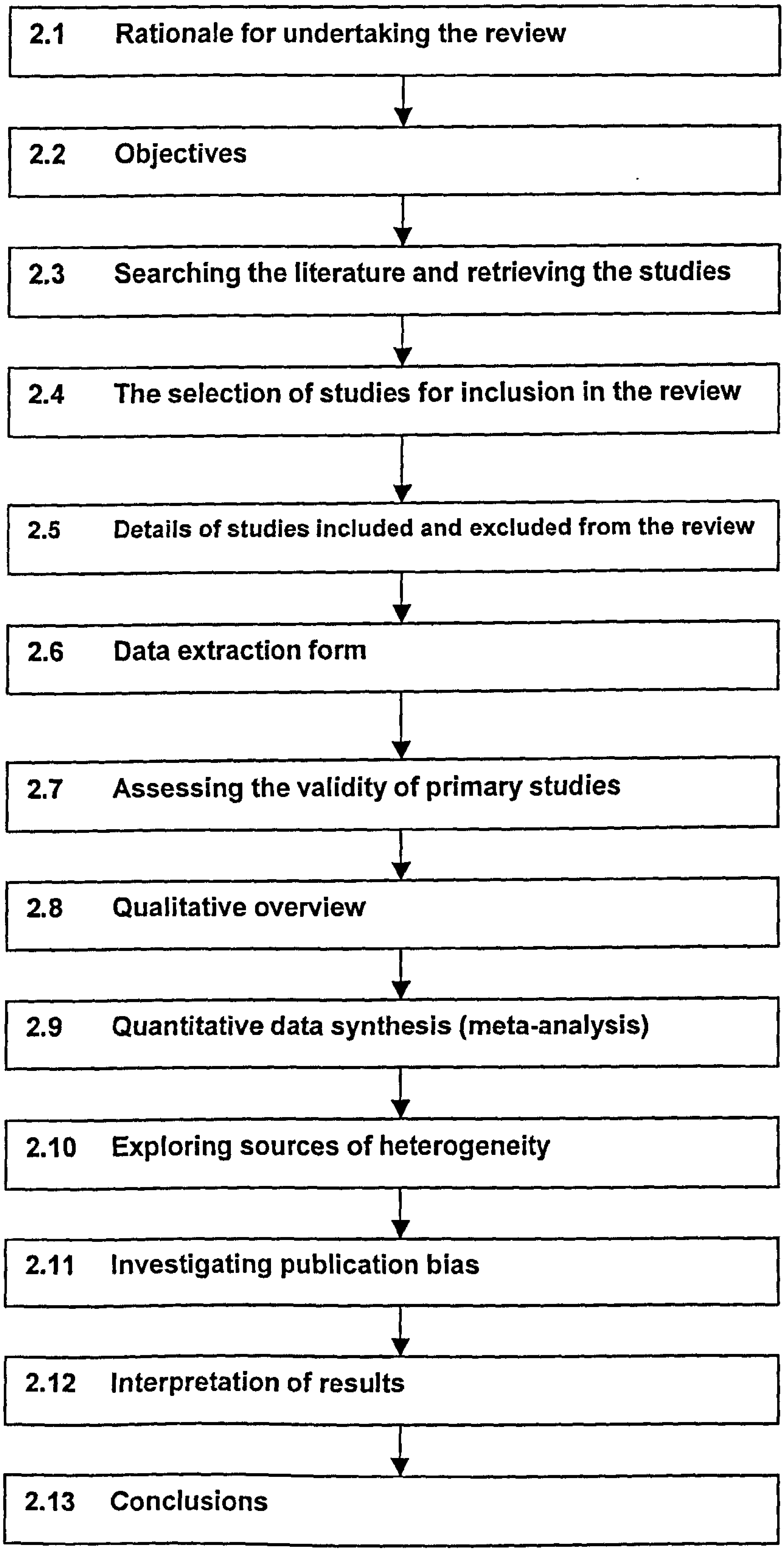
2.1.4.4 Framework for the review

Finally, when conducting the background research, various frameworks on how to conduct systematic reviews were discovered, including those under the auspices of the Centre for Reviews and Dissemination, University of York and the Cochrane Collaboration. These frameworks advocate that reviews should be designed to address explicit objectives, avoid biased selection of studies, accurately extract data, assess study validity, and interpret the results narratively or using statistical methods to produce a single estimate of a treatment effect (Sutton et al, 2000). The framework applied for this review is illustrated in Figure 2.2.

Having discussed the rationale for the review in section 2.1 the specific objectives of the review are delineated in section 2.2. Section 2.3 then describes the multiple methods used to locate studies. Explicit selection criteria as shown in section 2.4 were then developed to help decide which studies were eligible for inclusion. The results of these searches are presented in section 2.5. Section 2.6 then describes the process of developing the data extraction form so that data could be reliably extracted and the results of its subsequent application. Section 2.7 then discusses how the studies included in the review were assessed for validity. Section 2.8 provides a qualitative overview of the studies included in the review using the classification system described in section 2.1.4.3. This was followed by section 2.9 which applied the statistical methods to quantitatively synthesise the results of the studies included in the review. Then sections 2.10 and 2.11 respectively present an exploration into the sources of heterogeneity when quantitatively summarising the studies included in the review and an investigation into publication bias. Having described the studies included in the review (section 2.8), attempted quantitative synthesis (section 2.9), explored sources of heterogeneity (section 2.10) and investigated publication bias (section 2.11) it was then possible in section 2.12 to address all the objectives of the review in light of these findings. Finally, section 2.13 presents the conclusions drawn.

This systematic review therefore comprehensively addressed the methods of scientifically synthesising studies that assess radiographer plain radiograph reporting performance.

Figure 2.2 Framework for review about radiographer radiograph reporting performance



2.2 Objectives

2.2.1 Primary objectives

1. To assess the radiograph reporting performance of radiographers compared with a reference standard.

2.2.2 Secondary objectives

2. To compare the radiograph reporting performance of selectively trained radiographers with other health care professionals (e.g. radiologists).

3. To assess the radiograph reporting performance of selectively trained radiographers for different categories of patients (e.g. A&E, non-A&E).

4. To assess the radiograph reporting performance of selectively trained radiographers for different body areas (e.g. skeleton, chest and abdomen).

5. To assess the effectiveness of training programmes for improving radiographer radiograph reporting performance.

2.2.3 Supplementary objectives

6. To assess the radiograph reading performance of radiographers compared with a reference standard in the following roles:
    - Red dot role - placing a red dot on a radiograph when an abnormality is present
    - Triage – categorisation of radiographs as, for example, normal, abnormal or significantly abnormal.

7. To identify threats to the validity of studies that assess radiographer plain radiograph reading performance.

8. To identify evidence that demonstrates the clinical effects of radiographer plain radiograph reading performance on patient diagnosis, choice of management and the associated costs.

2.3 Searching the literature and retrieving the studies

Studies that are eventually published are likely to be a biased set, overestimating performance (Macaskill et al, 1995), so a comprehensive search employing multiple strategies was employed in an attempt to eliminate publication bias. Advice on developing and executing the search strategies was sought from an information specialist.

2.3.1 *Electronic databases*

The following electronic databases were searched. The strategies are in Annex 2.1:

- MEDLINE (*Index Medicus* online)
- Science Citation Index Expanded (SCI-Expanded)
- Cumulative Index to Nursing and Allied Health Literature (CINAHL)
- EMBASE (*Excerpta Medica* online)
- NHS National Research Register (NRR)
- Cochrane Library
- PsycINFO (*Psychological abstracts*) and
- System for Information on Grey Literature in Europe (SIGLE)

2.3.1.1 Searching MEDLINE

Several potentially eligible studies were located through MEDLINE by searching for known authors in the subject area. The index terms for the article (Medical Subject Headings [MeSH]) were identified to help develop the search strategy. The MeSH terms included: diagnostic-errors, sensitivity-and-specificity, observer-variation, fractures-radiography, radiography. Only the first two terms were used, as the others were accounted for in the remaining strategy or subsumed within these two. The title and abstract of the studies were also analysed, resulting in the inclusion of the following text words in the strategy: reporting, radiographs, radiographers, triage, x ray film(s). The explode facility was used for some text words, such as Radiography and Radiology, to search for narrower terms. Searching relevant papers for other synonyms was not found to be useful, but truncating the word 'radiographers' did help to improve the precision of the search.

2.3.1.2 Searching SCI-Expanded

Unlike using MEDLINE, when searching SCI-Expanded, it is only possible to enter individual search terms or phrases separated by search operators such as AND or OR. Only a single search statement was used, truncating the word 'radiographers', as this was considered the most appropriate word to help identify all potentially eligible studies.

2.3.1.3 Searching CINAHL

The CINAHL database is updated monthly and provides indices and abstracts of over 650 English language nursing and allied health journals, plus books and chapters. CINAHL was searched to locate papers published in allied health care journals. It is very similar to MEDLINE in the structure of records, the provision of Boolean commands, index terms and text word capabilities, and consequently the strategy employed for MEDLINE was replicated.

2.3.1.4 Searching EMBASE

EMBASE, like other databases, uses its own controlled vocabulary so many of the index terms used by the MEDLINE strategy were not applicable. A similar approach as described for searching MEDLINE was applied and identified the following useful index terms: observer-variation, error, diagnosis, diagnostic-imaging, diagnostic-accuracy. The same text words used for the MEDLINE strategy were included.

2.3.1.5 Searching NHS NRR

The NRR is a network of registers that record details of research and development projects taking place in or of interest to the NHS. Only the truncated word 'radiographer' was used, as in the SCI-Expanded search.

2.3.1.6 Searching the Cochrane Library

The Cochrane Library comprises several databases that generally include controlled trials or systematic reviews. The following were searched:

- Cochrane Database of Systematic Reviews (CDSR), which contains the full text of the regularly updated systematic reviews of the effects of health care prepared by The Cochrane Collaboration.
- Database of Abstracts of Reviews of Effects (DARE), which provides information on published reviews of the effects of health care.
- Cochrane Controlled Trials Register (CCTR), which is a list of references to controlled trials in health care.
- Heath Technology Assessment (HTA) database, which contains records of ongoing projects and publications reporting completed assessments of health care technologies.
- NHS Economic Evaluation database (NHS EED), which is a register of published economic evaluations of health care interventions.

To develop the search strategy, an eligible study was located using the truncated version of the word 'radiographers', which identified the following primary and secondary keywords: diagnostic-errors, radiography, radiology, observer-variation, triage. Using the thesaurus, the MeSH terms that contain these key words were identified. A combination of these MeSH terms and free text words were used to search the databases. When searching with the MeSH terms the explode facility in the thesaurus was used.

2.3.1.7 Searching PsycINFO

PsycINFO is an electronic version of *Psychological abstracts*, the leading abstracting and indexing publication for psychology and related disciplines. The relevance of psychological themes in image interpretation made it an appropriate database to search, but a new strategy was developed, as the usual index terms were not applicable. The key words were "cognitive-processes", for which the explode facility was used in the thesaurus, and "pattern" and "visual", for which the index and thesaurus were searched to identify other relevant terms: pattern-discrimination, pattern-identification, pattern-perception, pattern-recognition, visual-acuity, visual-perception, visual-search, visual-strategy, visual-tracking.

2.3.1.8 Searching SIGLE

SIGLE, supplied by WinSPIRS, is a bibliographic database covering European grey literature, such as reports, dissertations, books and conference proceedings in fields including economics and social sciences. Only the truncated version of the word 'radiographer' was used.

2.3.2 *Handsearching*

To underpin the electronic searches, the following list of journals and supplements was handsearched from 1990 onwards, to coincide with the acceleration of the debate following the NHS Community Care Act (1990): British Journal of Radiology (BJR); Clinical Radiology; Radiography Today/Synergy and Radiography (1995 onwards).

2.3.3 *Personal Communication*

Personal communication helps locate studies possibly missed by the electronic searches to avoid publication bias. The Royal College of Radiologists (RCR) and Society & College of Radiographers were contacted, resulting in communication with members of SIGRR and the following universities that provide postgraduate training in image interpretation: Bradford; Canterbury; Hertfordshire; Lancashire; Salford; South Bank.

2.3.4 *Reading reference lists*

The reference lists of all papers and reports identified using the preceding strategies were read for further potentially eligible studies.

2.3.5 *Grey Literature*

'Grey literature' comprises studies that are unpublished, have limited distribution and/or are not included in bibliographical retrieval systems (Last, 1995). It includes abstracts, unpublished studies, conference proceedings, graduate theses, book chapters, reports and applications (McAuley et al, 2000). Searching SIGLE, handsearching the BJR annual congress supplements, and personal communication were methods of locating grey literature. Identifying such studies was important, as their exclusion could exaggerate estimates of effectiveness (McAuley et al, 2000).

2.3.6 *Search limits*

Swinburne (1971) was the first to propose using radiographers to distinguish normal from abnormal radiographs, so when possible, databases were searched from 1971 to the end of October 2002. No language or geographical restrictions were applied.

2.4 The selection of studies for inclusion in the review

To minimise 'reviewer bias' the author and radiographer reviewer, using the title and abstract if provided, made an independent selection of eligible studies from the electronic databases. Full copies of articles were acquired if necessary. Complete agreement was found in the application of the criteria between the two reviewers.

For a study to be *included,* the following criteria should be met:

- Radiographer(s) were compared with a reference standard to assess their plain radiograph reading performance.
- Must include or have the potential to calculate an appropriate statistic that reflects accuracy (e.g. sensitivity, specificity).

The following criteria were used to *exclude* studies:

- Only included images from other modalities (e.g. mammograms, ultrasound scans)
- The study was not performed during 1971-2002/10
- Case reports of a radiographer reading radiographs for one or two patients
- Visual search strategy studies that used remote eye movement detection equipment to record visual search behaviour from a fixed distance and/or used phantom images. This is because of the unrealistic setting in which they were performed and focus on pattern recognition.

Finally, some studies were duplicated in different journals. Data from one study were sometimes incorporated in another. Since multiple publications based on the same data are a source of bias (Gotzsche, 1989), when there was evidence of duplication the following criteria were applied:

- when studies were re-published only the original paper was included
- abstracts later published as papers were excluded
- when studies re-used some data only the latest study was included.

2.5 Details of studies included and excluded from the review

2.5.1 Studies included in the review

This section describes the results of the search strategies.

2.5.1.1 Electronic searches

As seen in Table 2.2, 952 studies were retrieved from all electronic databases. On applying the selection criteria 25 remained, but several of these studies appeared in more than one database. Table 2.3 illustrates this overlap and shows that there were only eleven individual studies. The studies that correspond with the reference numbers are in the list of references in Annex 2.1. This table shows the importance of searching several databases as no single database identified all eleven eligible studies.

Table 2.2 Number of studies located from each database

Search resource	No of studies before selection criteria applied	No of studies after selection criteria applied
MEDLINE	25	6
SCI-Expanded	255	6
CINAHL	15	3
EMBASE	126	6
NRR	95	3
Cochrane Library	355	1
PsycINFO	58	0
SIGLE	23	0
Any database	952	25

Table 2.3 Overlap between databases for eligible studies

Ref. No	MEDLINE	SCI-E	CINAHL	EMBASE	NRR	Cochrane Library
29	✓	✓				
13	✓	✓				
5	✓	✓		✓		
16	✓	✓		✓		
7	✓	✓		✓		✓
6	✓	✓	✓	✓		
30			✓	✓		
2			✓	✓		
22					✓	
31					✓	
32					✓	

2.5.1.2 Handsearching

Handsearching identified nineteen new studies, of which thirteen were eligible abstracts located from searching conference supplements. The abstract by Webster & Gallacher (1998) comprised two separate studies, referred to as 26a and 26b in the results and discussion section, or 26 in the list of references in Annexe 2.1. Therefore, fourteen more studies were included.

2.5.1.3 Personal communication

Of the six university centres contacted, only the University of Hertfordshire did not provide data that could be included in the review. The universities providing data were Bradford, Salford, South Bank, Lancaster and Canterbury Christ Church University College. This strategy helped to locate a further ten eligible studies.

2.5.1.4 Grey Literature

Searching SIGLE did not identify any studies that satisfied the inclusion criteria. However, the fourteen studies identified by handsearching conference supplements were all abstracts and considered grey literature as defined in section 2.3.5. The ten further studies identified through personal communication were the results of audit, postgraduate training courses, and dissertations. In total 24/35 (69 per cent) studies could be defined as grey literature, which further emphasises the need for multiple methods of searching.

2.5.1.5 Reading reference lists

This approach did not identify any new studies.

2.5.2 Studies excluded from the review

Table 2.4 presents studies that involved radiographers reading radiographs but were excluded. These are listed in the references in Annex 2.1. The remainder of the review will focus on the included studies.

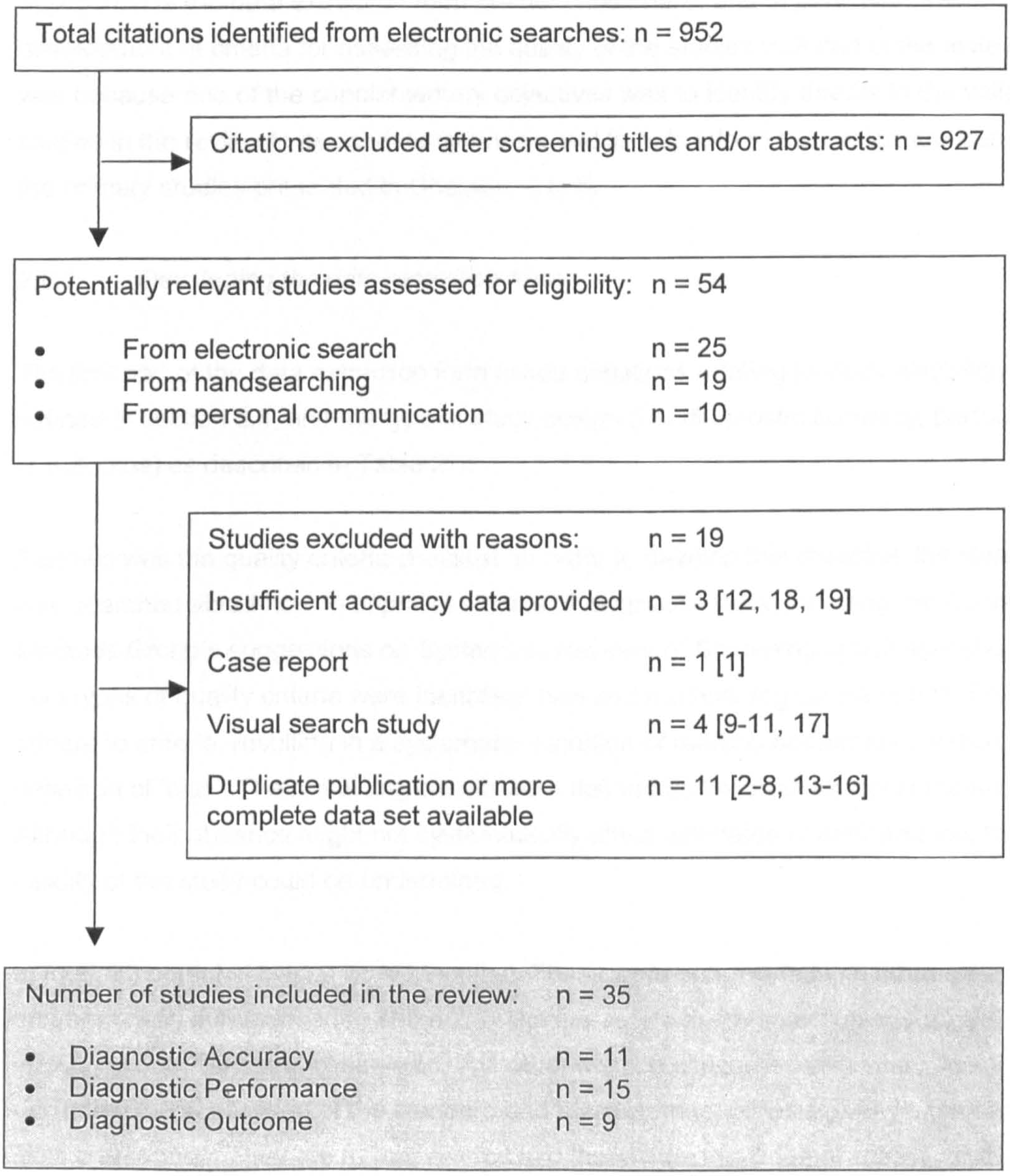
Table 2.4 Studies excluded from the review

Criteria	**Reference Number**
• Insufficient accuracy data provided	12, 18-19
• Case report study	1
• Visual search study	9-11, 17
• Duplicate publications or more complete data sets are available	2-8, 13-16

2.5.3 *Results of study selection process*

Figure 2.3 summarises the results of the process of selecting studies for inclusion in the review. The numbers in square brackets refer to the studies excluded from the review.

Figure 2.3 Study selection process for systematic review



2.6 Data extraction form

Having identified the 35 studies, it was necessary to develop the data extraction form presented in Annex 2.2. The form was to be used for recording study eligibility and design, assessing the quality of the studies, and recording factual information about, for example, the health care professionals being assessed and which type of patients and body areas were included in the sample of radiographs. The development, piloting, and results of the application of the data extraction form are described here, and in particular, the development of criteria for assessing the quality of the studies included in the review. This was because one of the supplementary objectives was to identify threats to the validity of studies in the review to assist data synthesis and to help inform the design and conduct of the primary studies presented in Chapters 4 to 6.

2.6.1 Developing the data extraction form

The first part of the data extraction form asked questions relating to study eligibility, as defined in section 2.4, and the type of study design (i.e. diagnostic accuracy, performance, or outcome) as described in Table 2.1.

Part two was the quality criteria checklist. In order to develop this checklist, the literature was searched about how to appraise studies of diagnostic tests, including the Cochrane Methods Group's suggestions on Systematic Reviews of Screening and Diagnostic Tests. Two types of quality criteria were identified: bias and methodological standards. Failure to adhere to criteria, resulting in a systematic distortion of reading performance indicated a definition of 'bias'. The remaining criteria were defined as 'methodological standards'. Although their absence might not systematically affect estimates of performance, the validity of the study could be undermined.

In total, 32 potential biases were identified. These were sub-divided into three categories and are briefly summarised in Table 2.5. The first refers to the selection of subjects (including both films and observers). The other two are concerned with study design (including the application of the standard and measurement of results) and interpretation of films and reports. They are further divided into those affecting internal validity, or the validity of the results within the context of the study, and those affecting external validity, or the generalisability of the results to other settings and populations.

Table 2.5 *Potential biases in radiograph reading performance studies*

Subjects [external validity]

Film selection biases
Referral bias
Referral (the time in the referral process that films are reported affects the case-mix of films)
Centripetal (the clinical setting can influence the prevalence and type of disease within the case-mix of films)
Popularity (experts may preferentially include and keep track of challenging or interesting cases)
Diagnostic access (geographical and financial factors affects access to the technology that produces the films)
Film selection (observers not reporting all the films eligible for inclusion)
Film filtering (no record of the criteria used to determine which films were eligible)

Film cohort bias
Spectrum (selecting films based on criteria such as prevalence of disease, severity and range of disease type)
Population (selecting a consecutive series of films over a suitable time period or a valid random sample)

Observer selection biases
Observer cohort (appropriate selection of observers)
Observer cohort comparator (appropriate matching of two or more groups of observers)

Study [internal validity]

Application of the gold standard
Verification (when not all films are read by the reference standard e.g. due to economic limitations)
Work-up (when not all films are read by the reference standard owing to the report of the observer)
Incorporation (an observers report is incorporated into the process of generating the standard)

Measurement of results
Disease progression bias (long delay between observers report and patient re-attendance as the standard)

Withdrawal bias
Loss to follow-up (films reported by the observer are lost and so the standard cannot be applied)
Indeterminate results (failure to include equivocal film interpretations)

Observer variability bias
Inter-observer variability (observers within a group independently read a sub-sample of the same films)
Intra-observer variability (the same observers independently read a sub-sample of films at a later date)

Arbiter variability bias
Inter-arbiter (a sub-sample of reports compared by independent arbiters)
Intra-arbiter (a sub-sample of reports compared by the same arbiter at a later date)

Interpretation [internal validity]

Independence of interpretations
Observer review (observers reporting films blind to the reference standard report)
Reference standard review (reference standard reporting films blind to the observers report)
Observer (observers reporting films independently)
Observer comparator (observers reading the same or a similar set of films)
Co-image (observers only having access to types of examination they are being evaluated to read)
Arbiter review (the arbiter also being an observer or reference standard)
Arbiter (the arbiter being blind to whether a report is made by an observer or reference standard)
Film access (the arbiter not having access to the films when comparing reports for concordance)
Clinical review (access to clinical details affecting observers performance)
Cohort comparator (both groups of observers reading films independently)
Co-image comparator (both groups of observers have similar access to films)
Arbiter comparator (the arbiter being blind to which reports belong to different groups of observers)

Ten methodological standards were also identified and subdivided into three groups: the selection of subjects (or films); study design; presentation of results. Table 2.6 lists the ten questions asked when assessing whether methodological standards had been met in plain radiograph reading performance studies.

Table 2.6 Methodological standards in radiograph reading performance studies

---

Selection of subjects (films)

1. Was an appropriate sample size considered?

Study design

2. Was a normal/abnormal report adequately defined?
3. Was the performance of the observers placed in the context of the diagnostic sequence?
4. Was the contribution of individual groups determined if the combined performance of two (or more) different groups of observers were assessed?
5. Was an appropriate (valid) reference standard ("gold" or "criterion") used?
6. Was an appropriate (valid) arbiter used to compare radiographers' reports with the reference standard?
7. Was an appropriate control used?

Presentation of results

9. Were films appropriately analysed for pertinent subgroups?
8. Was the data presented in enough detail to allow for the re-calculation of performance statistics (e.g. sensitivity and specificity) and confidence intervals?
10. Were indeterminate (i.e. equivocal; missing data, non-diagnostic) results appropriately presented?

Part two of the data extraction form comprised two major sections. The first asked thirty questions about bias in the selection of subjects and the conduct of the study. Each criterion in this section was recorded as "DONE", "NOT CLEAR", "NOT DONE", or "N/A" i.e. not applicable. The second section asked ten questions relating to the studies' adherence to methodological standards. In this section, the same options were available except for "NOT CLEAR", as these criteria were easier to judge for adherence.

Part three was developed using the Cochrane Effective Practice and Organisation of Care Group data extraction form. Factual information was recorded about the characteristics of the professionals and participants included in a study, the accuracy of its results including raw data, and whether it had addressed any clinical effects and costs of radiographers reading plain radiographs.

2.6.2 Piloting the data extraction form

Because so many questions were included in the data extraction form, it was important to test that the answers were reproducible and objective. The author and a researcher who was familiar with the subject area piloted the form by independently completing it for two eligible studies (7 and 16). The two studies chosen represented a different design in Table 2.1 and were comprehensively written up, making it possible to judge whether or not certain criteria had been met, rather than recording the UNCLEAR option because of insufficient information.

Table 2.7 illustrates the extent of agreement for the different sections of the form. For part one, there was some disagreement because the researcher was unfamiliar with the different types of study designs. Part two, sections one and two, showed 68 per cent and 75 per cent agreement respectively between the author and researcher and moderate Kappa scores of 0.54 and 0.59. Disagreements in part two were as a result of the researcher's unfamiliarity with some of the terminology or that unsubstantiated assumptions were made. There was 90 per cent agreement for part three. The form was judged sufficiently reproducible and simple to complete, considering that the researcher received no training but still achieved moderate agreement with the author. It was also concluded that owing to the high degree of agreement only the author need extract data to complete part three of the form.

Table 2.7 Pilot application of data extraction form

Checklist section	% agreement	Kappa (95% CI)
Part 1	64	
Part 2		
Section 1	68	0.54 (0.36, 0.72)
Section 2	75	0.59 (0.28, 0.90)
Part 3	90	
Total	73	

2.6.3 Applying the data extraction form

Paper forms were used to extract data, with the version number recorded to reduce the chance of erroneously using an outdated form. Each study was given a unique reference number (as shown in the list of references in Annex 2.1) and Microsoft Excel and Access were used to store and manage data. The author independently assessed all eligible studies. Eight studies required completion by personal communication, which involved visiting the investigator responsible or applying the form by telephone.

Table 2.8 illustrates which reviewers appraised the remaining studies. All three reviewers were familiar with many of the studies so there was no blinding to authors' names, institutions, journal of publication, or results. Discordance between reviewers was resolved by discussion, with the decisions recorded. The radiographer and radiologist reviewers also completed data extraction forms for their own studies. Rather than being a source of bias, this should improve the accuracy with which the form was completed when aspects of study design in a paper were not explicit.

Table 2.8 Reviewer(s) responsible for study appraisal

Reviewer	Reference Number
Radiographer	1, 3, 4, 6, 10, 13, 17, 18, 22, 23, 28-34
Radiologist	2, 5, 7, 8, 11, 16, 20, 24, 25, 27
Author	All studies

The extent of agreement between reviewers when applying part two, or the quality criteria component of the form, is presented in Table 2.9. For section one, the author compared with the radiographer and radiologist reviewer respectively had 82 per cent and 71 per cent agreement and good or moderate Kappa scores of 0.73 and 0.58. For section two, the author compared with the radiographer and radiologist reviewer respectively had 92 per cent and 86 per cent agreement and good to very good Kappa scores of 0.88 and 0.79. The higher Kappa scores between the author and radiographer reviewer is probably explained by the greater number of studies that the latter reviewed resulting in increased familiarity in the application of the form. However, all Kappa scores ranged from at least moderate to very good agreement, evidence of reliable decision-making between all reviewers. The data extracted when applying part two of the form is presented in Annex 2.3.

Table 2.9 Final application of the quality criteria checklist

Checklist section		**Kappa (95% CI)**	**(%)**	**Kappa (95% CI)**	**(%)**
		Radiographer		**Radiologist**	
Part 2	Section 1	0.73 (0.66, 0.79)	(82)	0.58 (0.50, 0.65)	(71)
	Section 2	0.88 (0.77, 0.99)	(92)	0.79 (0.65, 0.93)	(86)

2.6.4 Conclusion

This section provided evidence of development of a comprehensive form that could provide reproducible answers and address the objectives of the review. In particular parts two and three would elicit the data to address the supplementary objectives and in turn inform the design of the primary studies in Chapters 4 to 6. However, part two of the form comprised forty questions in total. The next section discusses how data extracted with part two of the form was used to reflect the validity of the studies included in the review without having to comment on individual quality criteria.

2.7 Assessing the validity of primary studies

When performing the background research to this review (as described in section 2.1.4.3), relevant studies were found to vary in design according to their setting and whether radiographers' alone or other health care professionals' performance was assessed. As a result, studies were classified in three groups: diagnostic accuracy, performance, and outcome. Although classification of study by design can help reflect validity, it was useful to develop a banding system for a more explicit reflection of the strength of evidence provided by studies included in the review. This should facilitate the synthesis of results from the studies in the review.

First, a numerical scoring scheme was produced so that each study could be awarded a quality score and ranked in order of validity. This could be achieved in a variety of ways, including asking experts to agree on which quality criteria should be awarded most weight because their absence could more seriously undermine study validity (Mulrow et al, 1989). Within the resource constraints of this review, it was only feasible for the author and radiographer reviewer to judge the importance of different quality criteria. Both reviewers independently recorded the importance of each criterion as high (3 points), medium (2 points) or low (1 point). The two reviewers then discussed their recordings and resolved disagreements by discussion.

Four possible scoring schemes were developed to reflect differences in assumptions reviewers made on the importance of criteria. The following describes these scoring systems and Table A2.4.1 in Annex 2.4 lists how important the different criteria were for each scoring system:

- SS1 (consensus) = the two reviewers independently scored the different criteria then came to a consensus as to the score attributable to each criterion.
- SS2 (alternative consensus) = having come to a consensus as to the scores in SS1, the two reviewers discussed rational alternative scores for the different criteria, which mainly involved changing the consensus score when the two reviewers' independent decisions disagreed for SS1.
- SS3 = the scores allocated by the author only.
- SS4 = the scores allocated by the radiographer reviewer.

The different scoring systems were applied to each study by summing the scores awarded to each criterion. When a criterion was judged "NOT CLEAR" it was assumed "NOT DONE". A score of zero was given when a particular criterion was "N/A" or the bias was absent. The score of a study was divided by the total score for that study if all applicable criteria were judged "NOT DONE", and then multiplied by one hundred to produce a normalised score. Studies awarded the lowest score were the most valid. Because ordinal data was used, a study with a score of 80 was not necessarily twice as poor in quality as a study with a score of 40. Each study was ranked in order of its score for each system as shown in Table A2.4.2. The score of each study based on the four systems was then aggregated and divided by four to produce a mean score that incorporated all the different decisions and assumptions made. The studies were then ranked in order of their mean score as shown in Table A2.4.3.

Having ranked the studies according to their mean quality score, it was then possible to group them into different bands. The reviewers could have done this by subjectively judging how the mean quality scores for the different studies naturally grouped together, but a statistical method, considered more objective and valid, was used. Cluster analysis is a statistical approach to inform the natural grouping (or clusters) of studies. It is an exploratory data analysis technique that uses systematic methods for testing how data groups together. The kmeans cluster analysis method was chosen, using the STATA statistical package (StataCorp, 2001). This method assigns the mean score for a study to a group of studies whose mean is closest and based on that categorisation new group means are determined. The steps are continued until no mean study scores change groups. The analysis requires the number of clusters to be specified. Table 2.10 presents the results of this analysis, showing the range of ranks in which the studies were grouped. A banding system of only three groups (or clusters) was used to aid clear demarcation of the higher from the lower quality studies and facilitate summarising study quality when interpreting results.

Table 2.10 System for banding studies by quality

Rank	Band	Quality
1-7	A	High
8-27	B	Moderate
28-35	C	Low

2.8 Qualitative overview

Before potential methods of data synthesis were explored, it was thought useful to provide an overview of the results of the 35 studies included in the review using the classification system presented in section 2.1.4.3. The results from the individual studies could be incorporated into section 2.12 for the final interpretation of the results of the review.

The summary tables in this section present information on the facts, results and overall quality of the primary studies included, ordered within tables by their quality using the mean scores and the banding system. The letters in square brackets refer to the quality of evidence and the numbers in superscript to the studies providing this evidence. The abbreviations used in the tables are presented in the List of abbreviations.

2.8.1 Diagnostic accuracy studies

2.8.1.1 Diagnostic accuracy studies that only assess radiographers only

Table 2.11 shows the results of six studies that assessed radiographer radiograph reading accuracy. The quality of these studies varies considerably, ranging from [A] [18] with a score of 10.1 to [C] [2] with a score of 55.3.

The two studies of highest quality show that trained radiographers can report A&E radiographs of the skeleton between 94 per cent and 95 per cent accuracy. For the same body areas they could report a combination of out-patient, in-patient and general practitioner radiographs between 93 per cent and 94 per cent accuracy [A] [18, 17]. A study of moderate quality provides additional evidence that radiographers post-training can report radiographs of the appendicular skeleton at 95 per cent accuracy [B] [14]. A further study of radiographers reporting A&E radiographs of the skeleton showed that at follow-up (six to ten weeks post-training) radiographers reported at only 81 per cent accuracy [B] [13], although the results indicate that this was a statistically significant improvement compared with the initial assessment ($P<0.01$).

Evidence of radiographer's accuracy when reading chest radiographs was provided in two studies and one also included abdomen radiographs [A] [11] [C] [2]. One study showed that radiographers not trained to report achieved 84 per cent accuracy when reporting chest radiographs for various types of patient [A] [11]. The other study showed that training radiographers to triage chest and thoracic cage radiographs significantly improved their accuracy from 63 per cent to 74 per cent for A&E radiographs (P<0.01) and from 63 per cent to 86 per cent for in-patient and out-patient radiographs (P<0.01) [C] [2]. These two studies, compared with the other studies in the table, showed that radiographers appear to read chest radiographs less accurately than for the skeletal system. Furthermore, one of the studies showed that untrained radiographers report radiographs of the abdomen at only 62 per cent accuracy [A] [11].

In summary, the two high quality studies show that under controlled conditions trained radiographers can report radiographs of the skeleton for different types of patient between 93 per cent and 95 per cent accuracy [A] [18, 17]. However, radiographers report chest and abdomen plain radiographs at a lower level of accuracy [A] [11] [C] [2]. There was also evidence, from studies of varying quality, that training programmes are effective [A] [17] [B] [13] [C] [2].

2.8.1.2 Diagnostic accuracy studies that assess radiographers and radiologists

Table 2.12 gives the results of four studies assessing both radiographer and radiologist reporting accuracy [A] [22, 3] [B] [23, 27].

The highest quality study demonstrates that radiographers during training reported A&E radiographs for all body areas more accurately than untrained radiographers and radiologist trainees [A] [22], though it is not known whether this was statistically significant. Another study shows that senior radiographers without training report A&E radiographs for all body areas significantly more accurately than first year Radiology registrars (P<0.05), not significantly different from second year registrars (P=0.43) but significantly less accurately than those who had recently completed their fellowship (P<0.05) [A] [3]. A further study showed that radiographers post-training reported A&E radiographs of the skeleton with considerably more accuracy than second year Radiology registrars (P<0.02) and were not significantly different from the third year registrars [B] [27]. The final study shows that radiographers trained to report achieved a level of accuracy when reporting A&E radiographs of the skeleton similar to second year Radiology registrars (51per cent vs 52 per cent) [B] [23]. Although the consultant radiologists reported at 57 per cent accuracy, this was not significantly different from the other two groups (P>0.05). The reason for the low accuracy in this study is probably explained by the careful selection of a difficult case-mix of twenty radiographs.

Table 2.11 Diagnostic accuracy studies that assess radiographers only

Ref	Study Type	Grey Lit?	Training/ grade	Sample		Patient type	Body area	Results			Study quality	
				Observers	Films			Sens	Spec	Acc	Score	Band
[18]	Reporting	Yes	Post	Radiographer: n = 28	N = 5595	A&E	Skeleton	93%	95%	94%	10.1	A
						OP, IP, GP	Skeleton	93%	93%	93%		
[17]	Reporting	Yes	Post (1994)	Radiographer: n = 7	n = 100	A&E	Skeleton	95%	96%	95%	19.3	A
			Pre (1995)	Radiographer: n = 6	n = 100	A&E	Skeleton	90%	92%	91%		
			Post (1995)	Radiographer: n = 5	n ~ 100	A&E	Skeleton	95%	93%	94%		
			Post (1995)	Radiographer: n = 5	n ~ 100	OP, IP, GP	Skeleton	97%	94%	94%		
[11]	Reporting	Yes	Untrained	Radiographer: n = 73	n = 79	A&E, OP, IP, GP	Chest/Abdomen	91%	63%		30.6	A
							Chest			84%		
							Abdomen			62%		
[13]	Reporting	No	Pre (I)	Radiographer: n = 22	n = 42	A&E	Skeleton	92%	65%	71%	38.4	B
			Post (II)	Radiographer: n = 22	n = 42	A&E	Skeleton	100%	53%	65%		
			Follow-up (III)	Radiographer: n = 22	n = 42	A&E	Skeleton	96%	75%	81%		
								Friedman (for I, II & III)				
								P=0.038	P<0.001	P<0.001		
								Wilcoxin Rank Sum (I vs III)				
								P=0.197	P<0.002	P<0.005		
[14]	Reporting	Yes	Post (1995-98)	Radiographer: n = 22	n = 70	General sample	Append	96%	93%	95%	43.7	B
[2]	Triage	No	Pre	Radiographer: n = 25	N = 76	A&E	Chest, thoracic cage	88%	40%	63%	55.3	C
			Post	Radiographer: n = 26	N = 142	A&E	Chest, thoracic cage	100%	68%	74%		
								Hypothesis test (pre vs post)[a]				
								P=0.005	P=0.059	P=0.015		
			Pre	Radiographer: n = 25	N = 121	IP, OP	Chest, thoracic cage	86%	55%	63%		
			Post	Radiographer: n = 26	N = 342	IP, OP	Chest, thoracic cage	92%	84%	86%		
								Hypothesis test (pre vs post)[a]				
								P=0.317	P<0.001	P<0.001		

[a] Hypothesis test for comparing proportions in two independent groups (Altman, 1991 p.234)

Table 2.12 Diagnostic accuracy studies that assess radiographers and radiologists

Ref	Study Type	Grey Lit?	Training/ grade	Sample		Patient type	Body area	Results			Study quality	
				Observers	Films			Sens	Spec	Acc	Score	Band
[22]	Reporting	Yes	During	Radiographer: n = 4	n = 50	A&E	Skeleton	93%	97%	95%	24.6	A
					n = 50		Chest	91%	79%	83%		
					n = 40		Abdomen	100%	70%	78%		
			Untrained	Radiographer: n = 4	n = 50	A&E	Skeleton	92%	78%	84%		
					n = 50		Chest	84%	67%	72%		
					n = 40		Abdomen	100%	53%	65%		
			2nd yr regs	Radiologist: n = 4	n = 50	A&E	Skeleton	85%	78%	81%		
					n = 50		Chest	73%	72%	73%		
					n = 40		Abdomen	98%	45%	58%		
[3] [a]	Reporting	No	Untrained (R)	Radiographer: n = 7	n = 30	A&E	Skeleton, chest, abdomen			47.3	28.0	A
			1st yr regs (AI)	Radiologist: n = 7	n = 30	A&E	Skeleton, chest, abdomen			37.5		
			2nd yr regs (AII)	Radiologist: n = 7	n = 30	A&E	Skeleton, chest, abdomen			55.7		
			1st yr regs (B)	Radiologist: n = 7	n = 30	A&E	Skeleton, chest, abdomen			34.4		
			Post-fellows (C)	Radiologist: n = 6	n = 30	A&E	Skeleton, chest, abdomen			67		
[23] [b]	Reporting	Yes	Post	Radiographer: n = 27	n = 20	A&E	Skeleton			51%	36.8	B
			Consultants	Radiologist: n = 10	n = 20	A&E	Skeleton			57%		
			Trainees	Radiologist: n = 12	n = 20	A&E	Skeleton			52%		
[27] [c]	Reporting	Yes	Post (I)	Radiographer: n = 7	n = 45	A&E	Skeleton			78.14	37.0	B
			3rd yr regs (II)	Radiologist: n = 10	n = 45	A&E	Skeleton			72.50		
			2nd yr regs (III)	Radiologist: n = 8	n = 45	A&E	Skeleton			68.55		

[a] Films in this study were marked using a 1-3 point system. The following results are based on unpaired and paired Student t tests: AI vs B (P=NS); R vs AI (P<0.05); R vs B (P<0.05); R vs AII (P=0.43); AII vs AI (P<0.05); AII vs B (P<0.05); C vs R and C vs AII (P<0.05); C vs AI and C vs B (P<0.001).
[b] Although the consultant radiologists reported at 57% accuracy this was not significantly different to the other two groups of observers (p>0.05).
[c] Films in this study were marked using a specially developed scheme. The following results were based on a one-way ANOVA: I vs II (P=NS); I vs III (P<0.02); II vs III (P=NS).

2.8.1.3 Diagnostic accuracy studies that assess radiographers and A&E staff

Table 2.13 presents the one study that assessed radiographers and A&E staff reporting accuracy [A] [10] and demonstrated that trained radiographers reported A&E radiographs of the skeleton at 93 per cent accuracy. This was significantly more accurate than radiographers without training (P<0.01) and A&E nurses who did or did not receive training (P<0.01). The untrained radiographers also reported significantly more accurately than untrained A&E nurses (P=0.02), who tended to miss abnormalities, and were not significantly different from A&E nurses who had received training (P=0.67).

2.8.2 Diagnostic performance studies

Table 2.14 presents all fifteen diagnostic performance studies. No studies were banded A, eleven were banded B and four banded C.

Two studies, although of low quality, demonstrated that radiographers read radiographs in a red-dot role for all body areas or skeleton respectively at 90 per cent and 92 per cent accuracy [C] [25, 33]. A further study demonstrated that radiographers could red dot chest (including thoracic cage) radiographs for in-patients at a cardio-thoracic centre at 98 per cent accuracy [B] [29]. Another study assessed the ability of radiographers in a teaching hospital to triage A&E radiographs [B] [7], and with no additional training, triaged radiographs of the skeleton at 92 per cent accuracy. This was significantly more accurate than their triaging of chest/abdomen radiographs (P<0.001).

Six studies, of varying quality, also assessed trained radiographers reporting A&E radiographs of the skeleton [B] [28, 12, 15, 19] [C] [21, 34]. The highest quality of these studies assessed the implementation of radiographer reporting in four NHS Trusts [B] [28] and found that the radiographers reported A&E radiographs of the skeleton at 99 per cent accuracy. The results of the other five studies also showed that trained radiographers reported A&E radiographs of the skeleton between 94 per cent and 97 per cent accuracy.
Three studies also found that radiographers during training reported skeletal radiographs referred from A&E and other sources between 96 per cent and 98 per cent accuracy [B] [30, 9, 8]. Another showed that radiographers without training accurately reported 95 per cent of A&E radiographs of the skeleton [B] [32] and a further study found that radiographers with training reported A&E or GP radiographs of the appendicular skeleton at 97 per cent accuracy [B] [26a].

Table 2.13 Diagnostic accuracy studies that assess radiographers and A&E staff

Ref	Study Type	Grey Lit?	Training/ grade	Sample		Patient type	Body area	Results			Study quality	
				Observers	Films			Sens	Spec	Acc	Score	Band
[10]	Reporting	Yes	Post (A)	Radiographer: n = 22	n = 120	A&E	Skeleton	92%	93%	93%	25.8	A
			Untrained (B)	Radiographer: n = 10	n = 120	A&E	Skeleton	73%	86%	80%		
			Post (C)	A&E nurse: n = 4	n = 120	A&E	Skeleton	68%	92%	81%		
			Untrained (D)	A&E nurse: n = 5	n = 120	A&E	Skeleton	63%	86%	75%		
								Mann-Whitney				
								A vs B				
								P<0.001	P=0.007	P<0.001		
								A vs C				
								P=0.002	P=0.641	P=0.004		
								A vs D				
								P=0.001	P=0.009	P=0.001		
								B vs C				
								P=0.256	P=0.256	P=0.666		
								B vs D				
								P=0.074	P=0.854	P=0.019		

Table 2.14 Diagnostic performance studies that assess radiographers

Ref	Study Type	Grey Lit?	Training/ grade	Sample		Patient type	Body area	Results			Study quality	
				Observers	Films			Sens	Spec	Acc	Score	Band
[28]	Reporting	Yes	Post	Radiographer: n = 10	N = 7148	A&E	Skeleton	98%	99%	99%	33.2	B
[30]	Reporting	No	During	Radiographer: n = 1	N = 438	A&E	Append	97%	98%	98%	34.7	B
[12]	Reporting	Yes	Post	Radiographer: n = 14	N = 6980	A&E	Skeleton	92%	98%	96%	35.2	B
[26a]	Reporting	Yes	Post	Radiographer: n = 1	n = 1000	A&E, GP	Append	96%	98%	97%	37.4	B
[15]	Reporting	Yes	Post	Radiographer: n = 1	n = 790	A&E	Skeleton			97%	39.8	B
[29]	RDS	No	Post	Radiographer: n = 17	N = 8614	IP	Chest, thoracic cage	91%	99%	98%	40.7	B
[19]	Reporting	Yes	Post	Radiographer: n = 2	N = 5566	A&E	Skeleton			96%	40.8	B
[7][a]	Triage	No	Untrained	Radiographer: n = 40	N = 2844 N = 1025	A&E	Skeleton Chest/abdomen	89% 79%	93% 91%	92% 85%	41.0	B
[9]	Reporting	Yes	During	Radiographer: n = 3	N = 4999	A&E, orthopaedic	Skeleton	91%	98%	96%	42.4	B
[8]	Reporting	Yes	During	Radiographer: n = 5	N = 2999	A&E, OP	Append	97%	98%	98%	46.7	B
[32]	Reporting	Yes	Untrained	Radiographer: n = 21	N = 199	A&E	Skeleton	96%	95%	95%	48.2	B
[21]	Reporting	Yes	Post	Radiographer: n = 2	N = 1104	A&E	Append	93%	95%	94%	49.7	C
[34]	Reporting	Yes	Post	Radiographer: n = 32	N = 1021	A&E	Skeleton	90%	96%	95%	55.2	C
[25]	RDS	No	Untrained	Radiographer: n = 15	N = 3603	A&E	All areas	89%	73%	90%	61.8	C
[33]	RDS	Yes	Pre Post	Radiographer: n = 9	N = 493 N = 546	A&E	Skeleton	71% 81%	96% 96%	89% 92%	80.0	C

[a] The results exclude cases when radiographers required further advice

2.8.3 *Diagnostic outcome studies*

2.8.3.1 Diagnostic outcome studies that assess radiographers and radiologists

Table 2.15 presents results from four studies of varying quality that assessed radiographers and radiologists radiograph reporting performance. The highest quality study was [A] [26b] and the lowest was [C] [16].

The high quality study showed similar levels of accuracy for a radiographer, eight consultant radiologists, and six senior casualty officers when reporting A&E radiographs for the appendicular skeleton [A] [26b]. One of the two moderate quality studies found no significant difference between radiographers and radiologists of varying seniority when reporting A&E radiographs for all body areas ($P>0.2$) [B] [5], though in contrast, the other found that trained radiographers reported significantly more accurately than Radiology registrars ($P<0.001$) [B] [4]. This occurred despite the two groups in both studies reporting A&E radiographs for the same body areas, the most likely explanation being that the latter study only included a group of radiologist registrars who achieved 80 per cent sensitivity, missing abnormal films or 'under-calling' [B] [4].

The remaining study showed that during the first and second half of a training programme the radiographers reported A&E radiographs of the skeleton significantly more accurately, from 93 per cent to 97 per cent ($P<0.01$) [C] [16]. The difference in sensitivity at the start of the study between radiologist and radiographer was highly significant ($P<0.01$), but there was no statistically significant difference during the last two months ($P=0.19$). However, the difference in specificity between the radiologists and radiographers remained highly significant at the beginning and end of the study ($P<0.01$).

Table 2.15 Diagnostic outcome studies that assess radiographers and radiologists

Ref	Study Type	Grey Lit?	Training/ grade	Sample		Patient type	Body area	Results			Study quality	
				Observers	Films			Sens	Spec	Acc	Score	Band
[26b]	Reporting	Yes	Post	Radiographer: n = 1	N = 566	A&E	Append	96%	98%	97%[a]	31.7	A
			Consultants	Radiologist: n = 8	N = 566	A&E	Append	97%	99%	98%[a]		
			Senior	CO: n ~ 6	N = 566	A&E	Append	96%	95%	96%[a]		
[5]	Reporting	No	Post	Radiographer: n = 2	N = 561	A&E	Skeleton, chest, abdomen	92%	97%	96%	40.8	B
			Consultants, senior and 3rd year registrars	Radiologist: n = 12	N = 561	A&E	Skeleton, chest, abdomen	95%	98%	97%		
								McNemar's				
								P>0.2	P>0.2	P>0.2		
[4]	Reporting	Yes	Post	Radiographer: n = 2	N = 785	A&E	Skeleton, chest, abdomen	92%	100%	99%	41.8	B
			Registrars	Radiologist: n = ?	N = 785	A&E	Skeleton, chest, abdomen	80%	98%	96%		
								McNemar's				
								P<0.02	P<0.02	P<0.001		
[16]	Reporting	No	Start (1st 3m)	Radiographer: n = 3	N = 1946	A&E	Skeleton	83%	95%	93%	54.7	C
			End (2nd 3m)	Radiographer: n = 3	N = 1625	A&E	Skeleton	94%	97%	97%		
			During (all 6m)	Radiographer: n = 3	N = 3571	A&E	Skeleton	88%	96%	95%		
			Consultants	Radiologist: n = ?	N = 3571	A&E	Skeleton	97%	100%	99%[a]		
								Hypothesis test[b] (1st vs 2nd 3m)				
								P<0.001	P=0.006	P<0.001		

[a] This value is assumed based on the sensitivity and specificity scores
[b] Hypothesis test for comparing proportions in two independent groups (Altman, 1991 p.234)

2.8.3.2 Diagnostic outcome studies that assess radiographers and A&E staff

Table 2.16 gives results from five diagnostic outcome studies assessing radiographers and A&E staff plain radiograph reading performance. Three studies were banded [B] [31, 20, 6] and two banded [C] [24, 1].

Only the highest quality study assessed radiographers and casualty nurses without training in a reporting role. The radiographers and casualty nurses respectively reported A&E radiographs for the appendicular skeleton at 91 per cent and 83 per cent accuracy [B] [31].

Two studies provided moderate quality evidence about the performance of radiographers without training and casualty officers [B] [20, 6]. One assessed radiographers' performance in a red dot role and then a triage scheme [B] [20]. The triage scheme required them to attach a 'red dot' to abnormal radiographs, a 'blue dot' to normal radiographs, and no dot when they were uncertain about the presence of an abnormality or had no opportunity to dot a radiograph. For both schemes, the radiographers read A&E radiographs for all body areas more accurately than the casualty officers, but it was not known whether these findings were statistically significant. The other study also assessed radiographers and casualty officers in a red dot role when reading A&E radiographs for all body areas and both achieved the same accuracy, i.e. 93 per cent [B] [6].

Finally, two studies provide low quality evidence of radiographers performance combined with A&E staff [C] [24, 1]. Both studies found that the combined performance of the two professional groups was better than any single profession reporting radiographs alone. For instance, one study showed that when reading A&E radiographs of the skeleton, senior house officers alone, radiographers alone, and the two professional groups combined attained 93 per cent, 96 per cent and 98 per cent accuracy respectively [C] [24]. The other study found that the accuracy of radiographers triaging plus casualty nurses was 93 per cent - higher than the radiographers alone at 90 per cent accuracy [C] [1]. For both studies the results of significance tests were not provided.

2.8.4 Conclusion

This section described the 35 studies included in the review by study design. It was then necessary to judge which of these studies should be included in a meta-analysis to quantitatively address the primary and secondary objectives of the review.

Table 2.16 Diagnostic outcome studies that assess radiographers and A&E staff

Ref	Study Type	Grey Lit?	Training/ grade	Sample			Patient type	Body area	Results			Study quality	
				Observers		Films			Sens	Spec	Acc	Score	Band
[31]	Reporting	Yes	Untrained	Radiographer:	n = 40	N = 66	A&E	Append	95%	89%	91%	34.4	B
				ENP:	n = 10	N = 66	A&E	Append	89%	81%	83%		
[20]	RDS	Yes	Untrained	Radiographer:	n ~ 15	N ~ 3000	A&E	Skeleton, chest, abdomen	65%	98%	91%	42.2	B
				CO:	n ~ 12	N ~ 3000	A&E	Skeleton, chest, abdomen	78%	94%	85% [a]		
	Triage			Radiographer:	n ~ 15	N ~ 3000	A&E	Skeleton, chest, abdomen	92%	95%	94% [a]		
				CO:	n ~ 12	N ~ 3000	A&E	Skeleton, chest, abdomen	81%	95%	87% [a]		
[6] [b]	RDS	No	Untrained	Radiographer:	n = ?	N = 1496	A&E	Skeleton, chest, abdomen			93%	46.7	B
				CO:	n = ?	N = 1496	A&E	Skeleton, chest, abdomen			93%		
[24]	RDS	Yes	Untrained	Radiographer (R)	n = 35	N = 500	A&E	Skeleton			96%	53.5	C
				SHO:	n = 5	N = 500	A&E	Skeleton			93%		
				R + SHO:	n = 40	N = 500	A&E	Skeleton			98%		
[1]	Triage	Yes	Untrained	Radiographer (R)	n = 10	N = 406	A&E	Append (excl. pelvis)	87%	92%	90%	57.5	C
				R +ENP:	n = 13	N = 406	A&E	Append (excl. pelvis)	97%	89%	93%		

[a] These accuracy results exclude cases when observers declined to comment on radiographs
[b] The results exclude cases when observers were uncertain

2.9 Quantitative data synthesis (meta-analysis)

The rationale for the meta-analysis was to synthesise the evidence about radiographer performance when reporting plain radiographs. Section 2.9 was therefore divided into four sub-sections:

- the studies suitable for inclusion, to help inform the feasibility of meta-analysis;
- a summary of the statistical methods for producing pooled estimates of radiographer reporting performance to facilitate the choice of methods;
- application of the chosen methods to address the primary and secondary objectives about radiographer plain radiograph reporting;
- findings from the meta-analysis.

2.9.1 Is meta-analysis feasible?

Russell et al (1998) suggest that the following conditions should be met when deciding whether meta-analysis is feasible:

- there are sufficient data in an appropriate form;
- the settings and the study populations are generally homogeneous;
- the studies use outcome measures that are quantitative and generally compatible;
- the results are generally consistent, especially across levels of the design hierarchy.

The following exploration by study design helped develop the criteria used in judging studies for inclusion in the quantitative synthesis.

Diagnostic accuracy studies were explanatory in design and assessed the efficacy of radiographer reporting. The setting differed from the other study designs in that performance was assessed under controlled conditions. The case-mix of radiographs included a higher prevalence of abnormal cases, with a generally more robust reference standard. Irwig et al (1995) warn of the dangers of combining studies of diagnostic tests from different settings, so these were not combined with studies performed during clinical practice. There was also less clinical value in pooling the results of these studies so no meta-analysis was done using these data.

Diagnostic performance studies were conducted during clinical practice so were more pragmatic in design and assessed the effectiveness of radiographer performance. It was not appropriate to include studies that assessed radiographers in a red-dot or triage role as they differed from reporting studies. Eight of the remaining studies provided data in sufficient detail to construct 2x2 tables of radiographer reporting performance. Six studies were of moderate quality and two of low quality. Seven studies provided evidence of radiographers with some training reporting skeletal plain radiographs for different patient types. One study presented the results of untrained radiographers. The results of the eight studies showed that radiographer reporting performance was greater than 90 per cent accuracy. It was therefore judged reasonable to include all these studies in the meta-analysis.

Of the nine diagnostic outcome studies, only four provided data in sufficient detail to construct 2x2 tables to estimate radiographer reporting performance. Two studies were of moderate quality and two of low quality. These studies provided evidence of radiographers reporting A&E plain radiographs for a variety of body areas, with results again showing radiographer reporting performance at greater than 90 per cent accuracy. It was judged reasonable, on these grounds, to include these four studies in the meta-analysis.

In conclusion, the following criteria were used to judge whether studies should be included in the meta-analysis:

- The study must have been conducted in a clinical practice setting.
- The study must involve radiographers reporting radiographs.
- Data must be available to construct a 2x2 table so performance statistics like sensitivity and specificity could be calculated.

Results of the selected studies for inclusion in the meta-analysis are given in Figure 2.4. Table 2.17 presents the twelve studies eligible for inclusion in the meta-analysis in terms of study design, clinical factors, results and quality, ordered by quality. It was now necessary to discuss the different statistical methods available for producing a summary estimate of radiographers reporting performance before applying the chosen methods in practice.

Figure 2.4 *Study selection process for meta-analysis*

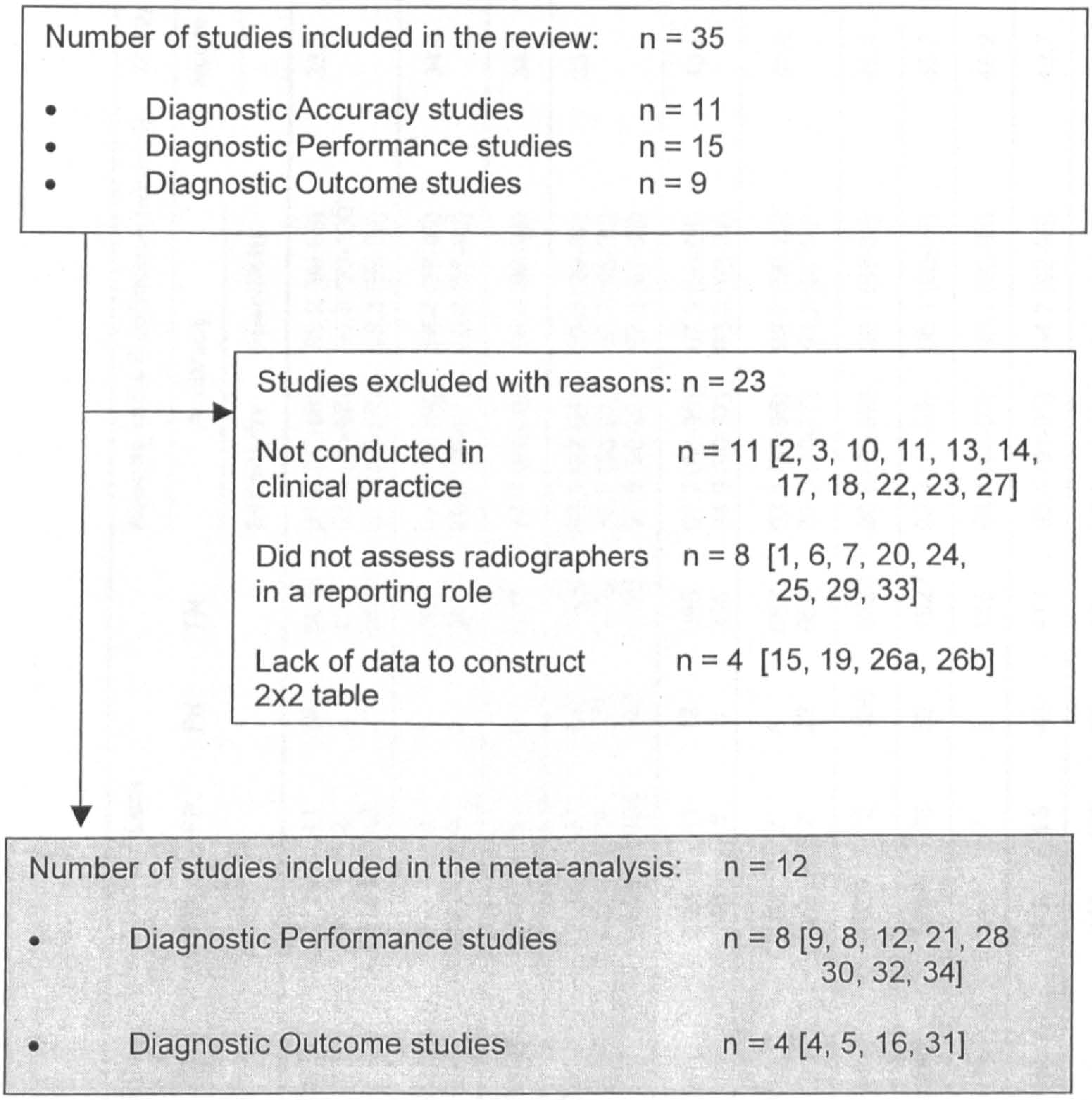


Table 2.17 Studies included in the meta-analyses

Ref	Profession	Study Design	Patient type	Body Area	Training	Data				Results (95% Confidence Interval)		Study quality	
						TP	FP	FN	TN	Accuracy		Score	Band
										Sensitivity	Specificity		
[28]	Radiographers	Performance	A&E	Append	Post	1164	41	29	5042	97.6 (97-98)	99.2 (99-99)	33.2	B
				Axial		31	2	2	837	93.9 (80-98)	99.8 (99-100)		
				Skeleton		1195	43	31	5879	97.5 (96-98)	99.3 (99-99)		
[31]	Radiographer	Outcome	A&E	Append	Untrained	18	5	1	42	94.7 (75-99)	89.4 (77-95)	34.4	B
	ENP					17	9	2	38	89.5 (69-97)	80.8 (67-90)		
[30]	Radiographers	Performance	A&E	Append	During	178	4	6	250	96.7 (93-99)	98.4 (96-99)	34.7	B
[12]	Radiographers	Performance	A&E	Append	Post	1112	67	82	1924	93.1 (92-94)	96.6 (96-97)	35.2	B
				Axial		920	37	106	2732	89.7 (88-91)	98.7 (98-99)		
				Skeleton		2032	104	188	4656	91.5 (90-93)	97.8 (97-98)		
[5]	Radiographers	Outcome	A&E	All body areas	Post	144	11	13	393	91.7 (86-95)	97.3 (95-98)	40.8	B
	Radiologists					149	8	8	396	94.9 (90-97)	98.0 (96-99)		
[4]	Radiographers	Outcome	A&E	All body areas	Post	108	2	8	667	93.1 (87-96)	99.7 (99-100)	41.8	B
	Radiologists					94	12	22	657	81.0 (73-87)	98.2 (97-99)		
[9]	Radiographers	Performance	A&E/orthop	Skeleton	During	1080	72	109	3738	90.8 (89-92)	98.1 (98-99)	42.4	B
[8]	Radiographers	Performance	A&E/OP	Append	During	925	78	69	1927	93.1 (91-95)	96.1 (95-97)	46.7	B
[32]	Radiographers	Performance	A&E	Skeleton	Untrained	54	7	2	136	96.4 (88-99)	95.1 (90-98)	48.2	B
[21]	Radiographers	Performance	A&E	Append	Post	618	23	46	417	93.1 (91-95)	94.7 (92-96)	49.7	C
[16]	Radiographers	Performance	A&E	Skeleton	During	440	129	61	2941	87.8 (85-90)	95.8 (95-96)	54.7	C
[34]	Radiographers	Outcome	A&E	Skeleton	Post	235	31	26	729	90.0 (86-93)	95.9 (94-97)	55.2	C

2.9.2 *Choice of statistical method*

The statistical methods used to pool results of studies of diagnostic accuracy depend on the choice of summary statistic, such as sensitivities and specificities, likelihood ratios (LR), and diagnostic odds ratios (DOR), Deeks (2001).

The pattern of heterogeneity observed between the results of the studies can also influence the choice of statistical method. Some variation in results is to be expected by chance, but variation might also be explained by factors such as type of patient or body area being reported and characteristics of the study design. One particular source of heterogeneity is changes in diagnostic threshold. Some studies may have done this explicitly, by varying the definition of a normal or abnormal report. For example, if considerable abnormality is required before a radiograph is declared positive, sensitivity will be low and specificity high; if less abnormality is required, sensitivity will be high and specificity low. There may also be naturally occurring variation in diagnostic thresholds between observers reading the radiographs and the arbiters responsible for comparing reports for concordance.

The possibility of a threshold effect can be investigated using receiver operating characteristic (ROC) curve methodology (Deeks, 2001). When such variation occurs the point estimates of studies presented graphically on a ROC plot (i.e. sensitivity *vs* 1-specificity) will reflect the curvature that parallels the underlying ROC curve, so rather than summarising the results of studies as a single estimate they can be summarised as a best fitting ROC curve. To help inform the choice of statistical method, the following discusses in more detail the meta-analytical techniques available.

2.9.2.1 Pooling sensitivities and specificities

This method should only be applied in the absence of variability in the diagnostic threshold. The threshold effect may be investigated both graphically and statistically. By plotting the study results on a ROC plot it is possible to visualise whether the points parallel a ROC curve and then undertaking tests of heterogeneity using standard chi-square tests, it can be seen whether differences among studies are due to chance alone. When estimates are homogenous, a summary point estimate of sensitivity and specificity and confidence intervals can be calculated, although when heterogeneity is present, it is necessary to explore the need to use ROC methodology.

2.9.2.2 Pooling likelihood ratios

Likelihood ratios (LRs) can also be pooled using standard methods of meta-analysis of risk ratios, which are algebraically identical to LRs (Simel et al, 1991). Heterogeneity of LRs can also be investigated using standard tests after combining the statistics in a meta-analysis. They are more informative than sensitivity and specificity alone, as LRs combine both (Chien & Khan, 2001).

2.9.2.3 Pooling DOR

The validity of producing summary estimates of DOR can also be investigated using standard methods of meta-analysis for combining odds ratios. The DerSimonian & Laird random effects model is recommended to reflect the heterogeneity often found in studies of diagnostic test accuracy (Laird & Mosteller, 1990; Berlin et al, 1989).

2.9.2.4 Pooling using ROC methodology

Evidence that the diagnostic threshold varies between studies makes it more appropriate to summarise the results using a ROC curve rather than a single summary estimate. The choice of ROC methodology depends on whether the curve is symmetrical around the "sensitivity equals specificity" line (Deeks, 2001). When the DOR for studies is constant, regardless of the diagnostic threshold, this should result in symmetrical ROC curves and under these circumstances, it is suggested that standard meta-analysis is used to estimate the summary DOR. Using this summary estimate, a ROC curve can be produced using a standard equation (Deeks, 2001).

When the method of pooling DOR provides evidence of heterogeneity, the DOR may change with diagnostic threshold. Asymmetrical ROC curves occur when the DOR changes with the threshold. Littenberg & Moses (1993) proposed a method for fitting ROC curves that allow for variation in DOR with threshold using the following model:

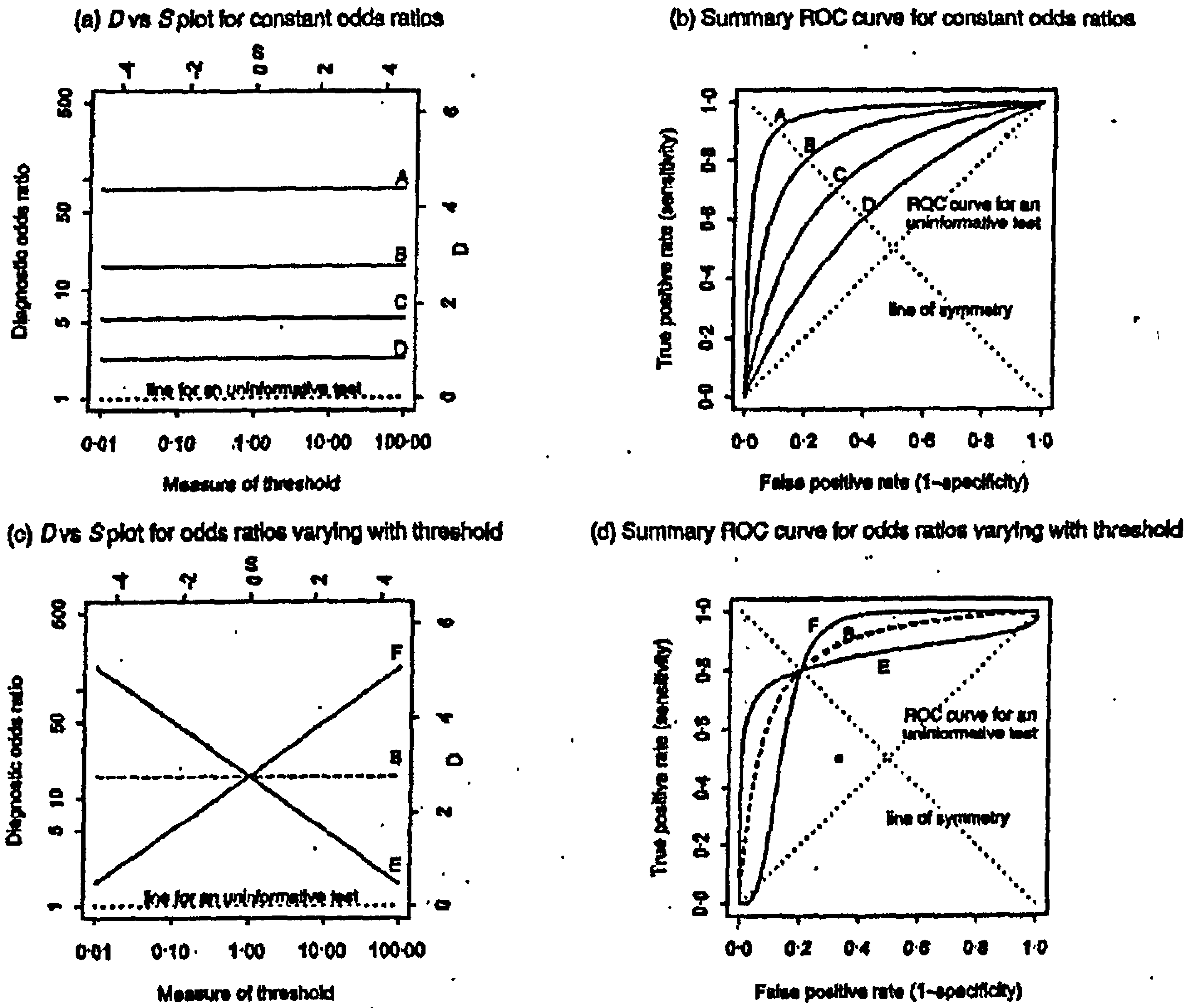
$$D = \alpha + \beta S$$

D is the natural logarithm of the DOR of the individual studies. *S* describes the leniency of the threshold for classifying a radiograph as abnormal. The intercept, α, is the estimated *D* when performance remains constant as the threshold varies from study to study, i.e. when sensitivity equals specificity ($S = 0$). This can be illustrated graphically by plotting *D* against *S*.

The regression coefficient, β, expresses variation in *D* across studies, providing an estimate of the extent to which *D* is independent of the threshold. When β is near zero, the shape of the curve calculated from the transformed model approximates that of a traditional ROC curve. Also, if β does not differ significantly from zero, performance does not depend on the threshold used in each study and studies can be summarised by a common DOR model (i.e. one where the DOR does not depend on threshold) given by the intercept alone. The larger the intercept, the closer the curve is positioned to the upper left corner in the ROC space and the greater the area under the ROC curve, which indicates greater performance. Figure 2.5 overleaf from Deeks (2001) shows D *vs* S plots and corresponding ROC curves with constant DOR (a) and (b) and DOR varying with diagnostic threshold (c) and (d).

This regression equation can be fitted using the equal weighted least squares method (EWLS) or the weighted method (WLS). The EWLS gives all the sample points equal weight, and does not emphasise the larger studies. To weight studies, the inverse variance of *D* is used which incorporates the sample size of each study, but a weighted analysis may bias the estimate. It is therefore recommended that both the weighted and unweighted results are presented (Sutton et al, 2000). Both methods also generally give similar estimates of α and β (Littenberg & Moses, 1993). To produce the ROC curve using this method it is possible to convert back from the D *vs* S plot to the conventional ROC plot axes by substituting the regression coefficient, β, and intercept, α, as shown in the formulas presented by Deeks (2001 p.274).

Figure 2.5 D-S plots and corresponding ROC curves with DOR



Source: Deeks (2001 p.274)

2.9.2.5 Conclusion

The studies included in this meta-analysis used sensitivity and specificity as measures of the accuracy of radiographer plain radiograph reporting performance, so the approach taken to address the primary objective was to first use a ROC plot that graphically presents whether point estimate of studies parallel an underlying ROC curve. Summary estimates of sensitivity and specificity were then calculated and tests of heterogeneity performed to see whether differences among studies were due to chance alone. Presence of heterogeneity indicated that a summary DOR should be computed to produce a best fitting ROC curve, assuming symmetry, then an asymmetrical ROC curve produced using Littenberg & Moses' (1993) method, which allows for variation in DOR with threshold. This would help to determine whether symmetry could be assumed and which ROC curve should be used as the summary measure of radiographer plain radiograph reporting performance.

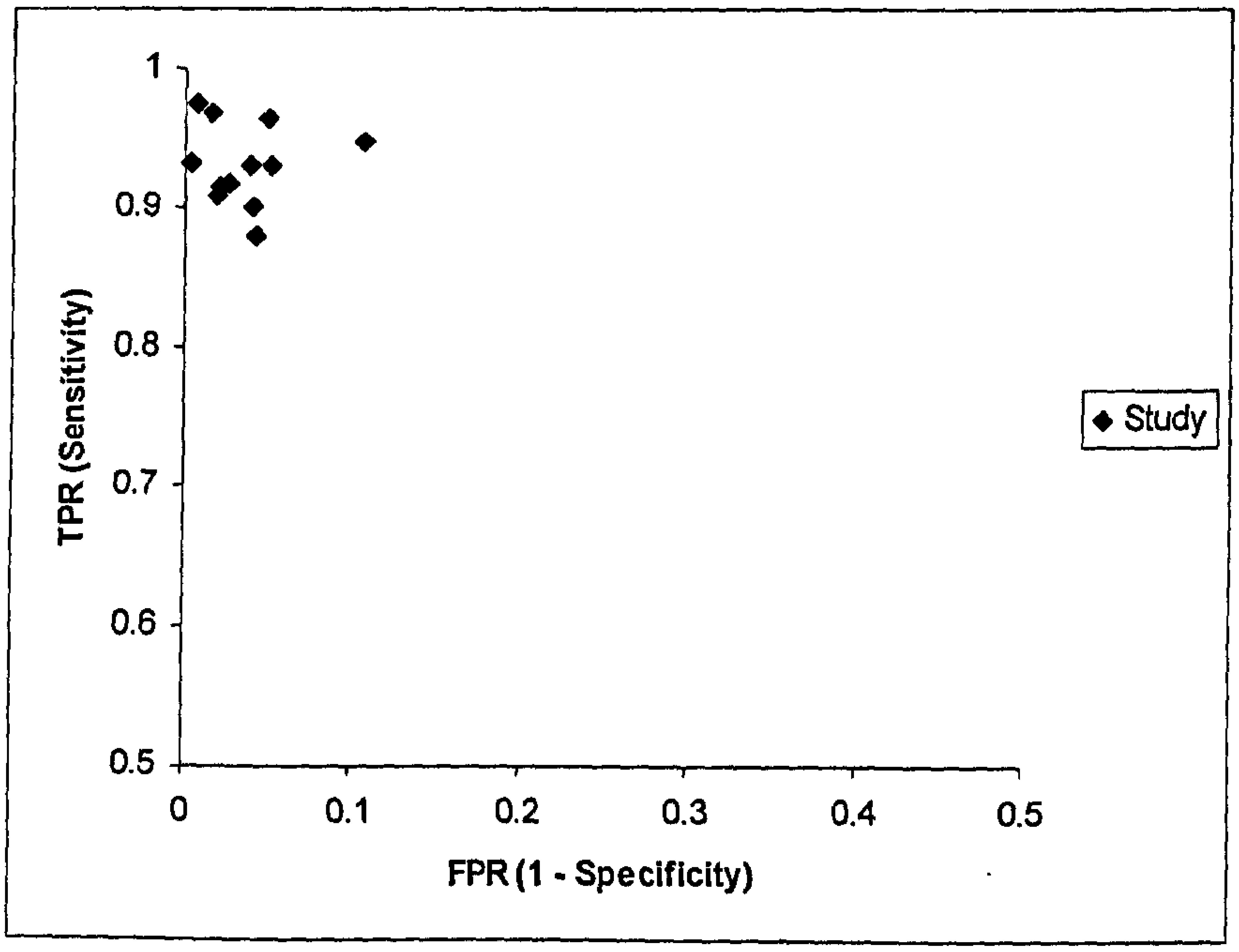
2.9.3 *Meta-analyses of radiographer plain radiograph reporting performance*

This section addresses the primary and secondary objectives of the thesis using meta-analytical techniques.

2.9.3.1 How accurately do radiographers report plain radiographs?

The primary objective was to assess the accuracy with which radiographers report plain radiographs. To determine whether the results of the twelve studies included in the meta-analysis could be summarised to address this objective, the point estimates of the studies were put on a ROC plot. Figure 2.6 illustrates the difficulty in judging whether the pattern of point estimates demonstrated curvature paralleling an underlying ROC curve, thus indicating variation in the diagnostic threshold between studies. Therefore, it was appropriate to calculate summary estimates of sensitivity and specificity and chi-squared tests to check homogeneity.

Figure 2.6 *ROC plot of radiographers plain radiograph reporting performance*



Figures 2.7 and 2.8 present the summary estimates of sensitivity and specificity, with the width of the 95 per cent CI reflecting the weight of a study, which are ordered as in Table 2.17. The summary sensitivity estimate of radiographer plain radiograph reporting performance was 92.6 per cent [95 per cent CI: (92.0, 93.2)] and specificity was 97.7 per cent [95 per cent CI: (97.5, 97.9)]. However, the chi-squared test confirms the statistical significance of the heterogeneity observed in the individual study estimates of sensitivity ($X^2 = 77.6$, df = 11, P<0.001) and specificity ($X^2 = 200.9$, df = 11, P<0.001) presented in the figures below. This highly significant heterogeneity made it appropriate to explore the use of ROC methodology as a method for summarising radiographer plain radiograph reporting performance.

Figure 2.7 Sensitivity estimates with 95 per cent CI

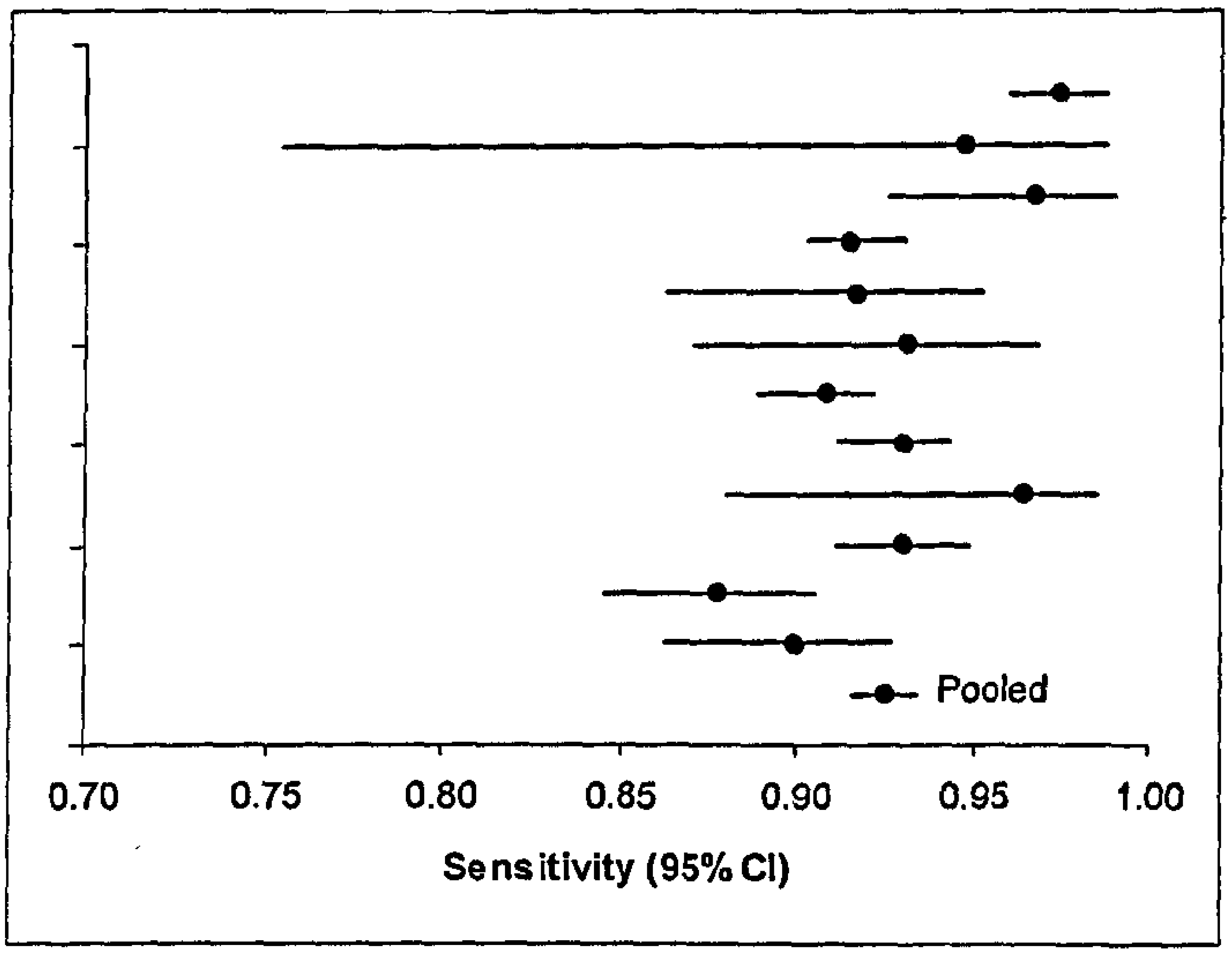
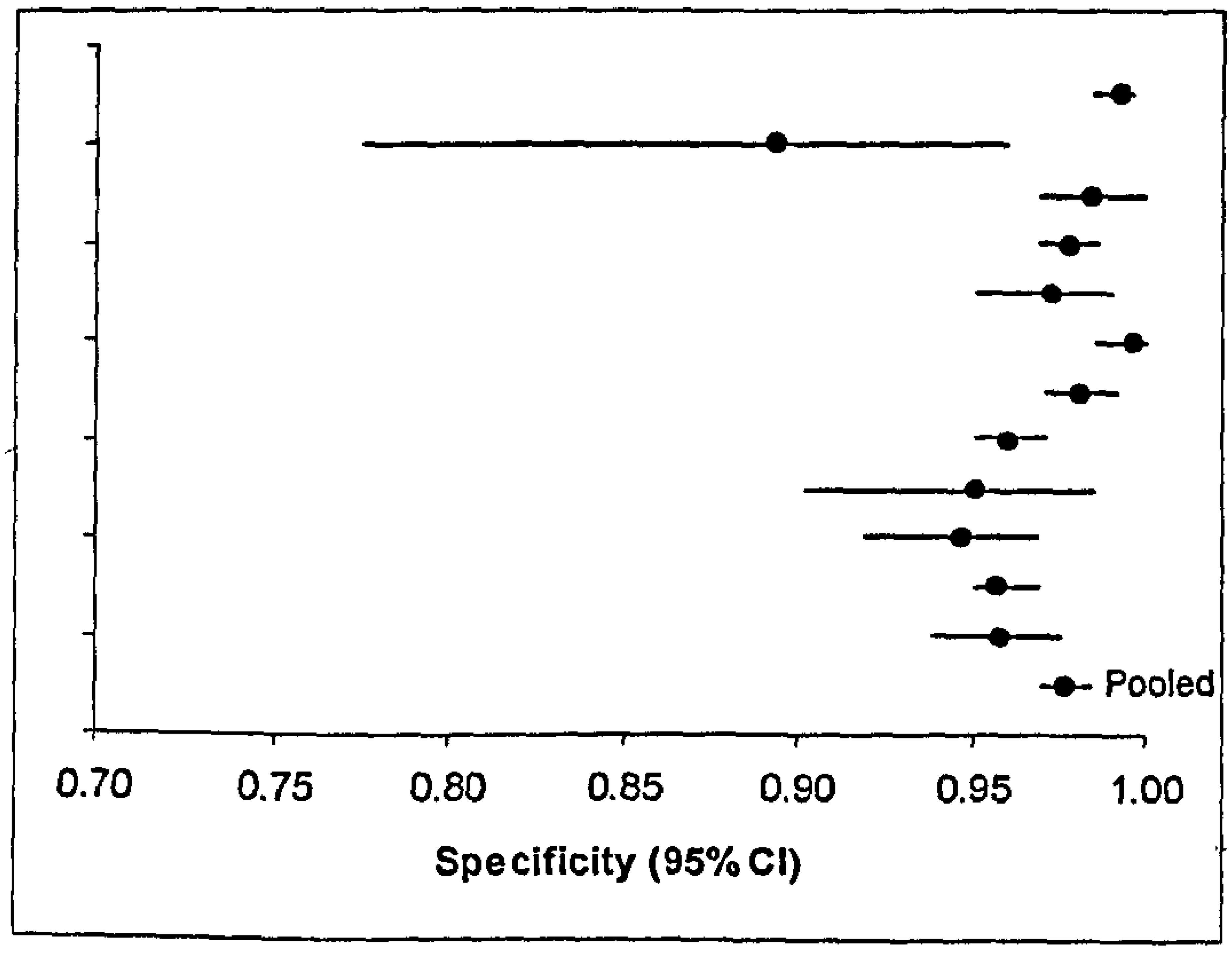


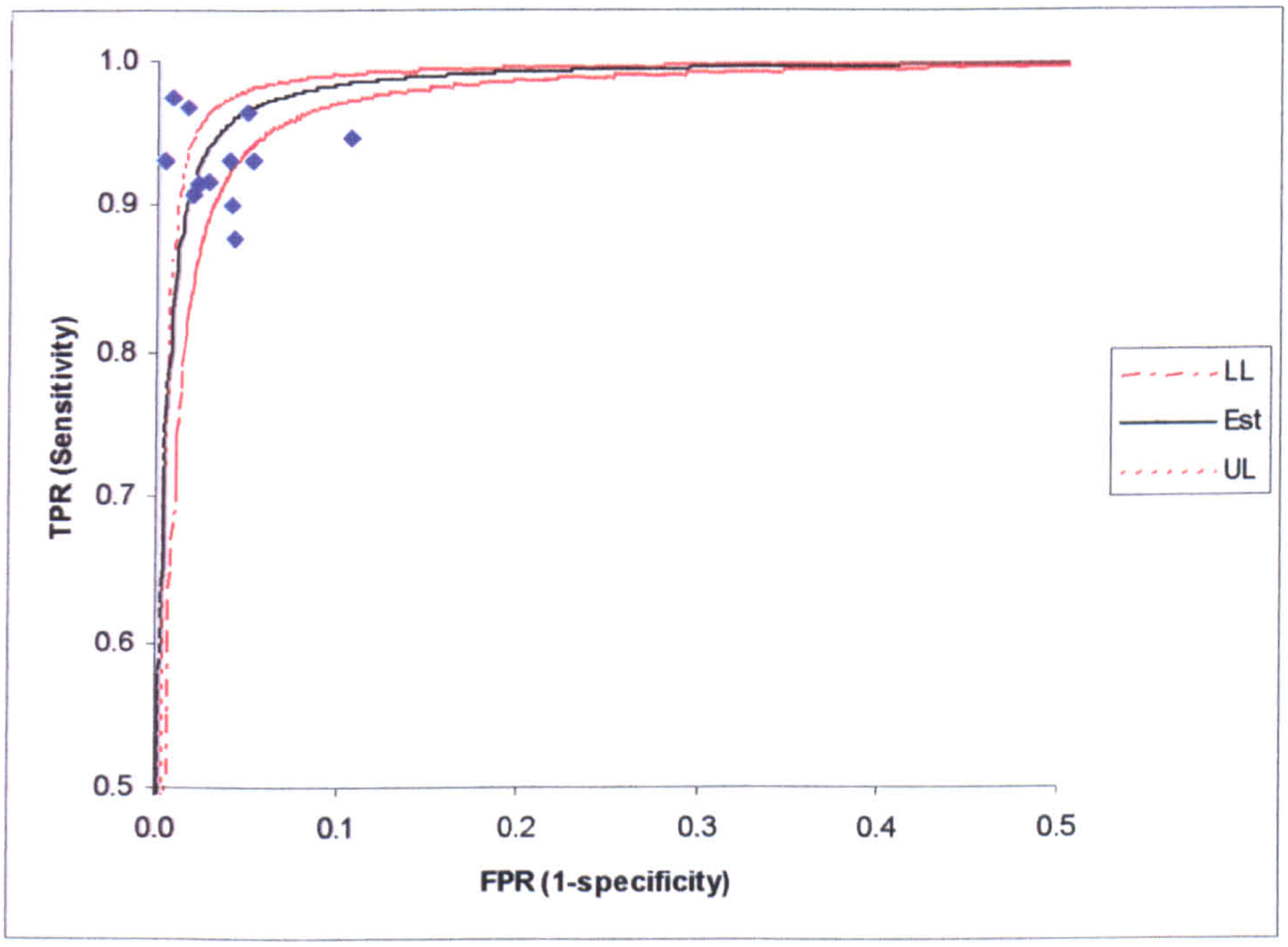
Figure 2.8 Specificity estimates with 95 per cent CI



Before using the Littenberg & Moses (1993) approach to test whether the curve was symmetrical around the "sensitivity equals specificity" line, it was assumed that DOR was constant regardless of the diagnostic threshold. The DerSimonian & Laird random effects model was used to calculate the DOR to be included in the equation, allowing the construction of the symmetric ROC curve with 95 per cent CI. This meta-analysis yielded a pooled DOR of 540.9 [95 per cent CI: (303.4, 965.3)] with evidence of heterogeneity (chi-square = 176.2, df = 11, P<0.00001).

Figure 2.9 illustrates the summary ROC curve for radiographers' plain radiograph reporting performance, assuming symmetry with 95 per cent CI to indicate precision of the estimated summary ROC curve (LL, lower limit; UL, upper limit). Methods of testing whether symmetry can be assumed were then applied.

Figure 2.9 Symmetric ROC curve of radiographers' radiograph reporting performance



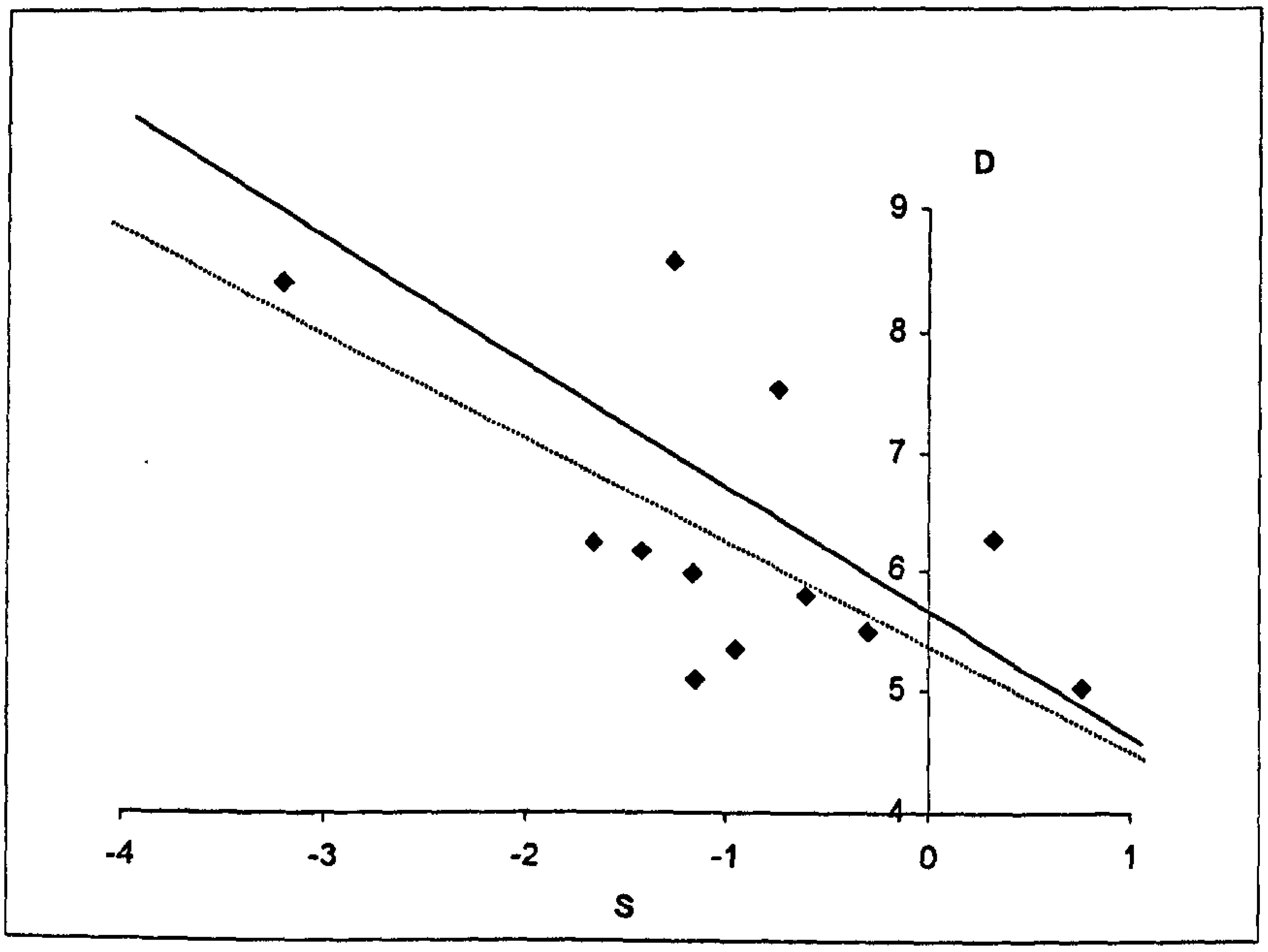
The Littenberg & Moses (1993) method was used to test whether the DOR, *D*, changes with diagnostic threshold for classifying a radiograph as abnormal, *S*. As illustrated in Table 2.18, the regression coefficient for *S* using EWLS was close to statistical significance at the 5 per cent level, but was not significant for WLS, although the wider confidence intervals for WLS reflect low precision. Both slope coefficients were similar for EWLS and WLS analyses respectively and β were not near zero, providing some evidence that *D* depends on the diagnostic threshold. Lines of best fit are displayed in Figure 2.10 (——, EWLS; ----, WLS).

The R^2 value, the square of the correlation coefficient between *D* and *S* and an indicator of the goodness-of-fit of the regression line, were low. This suggests limited correlation between *D* and *S*, but is possibly explained by the heterogeneity in plots in Figure 2.10. These findings suggest that variability in *D* between studies could be partly explained by the diagnostic threshold, although considerable variation was left unexplained.

Table 2.18 DOR based on a single threshold in each primary study

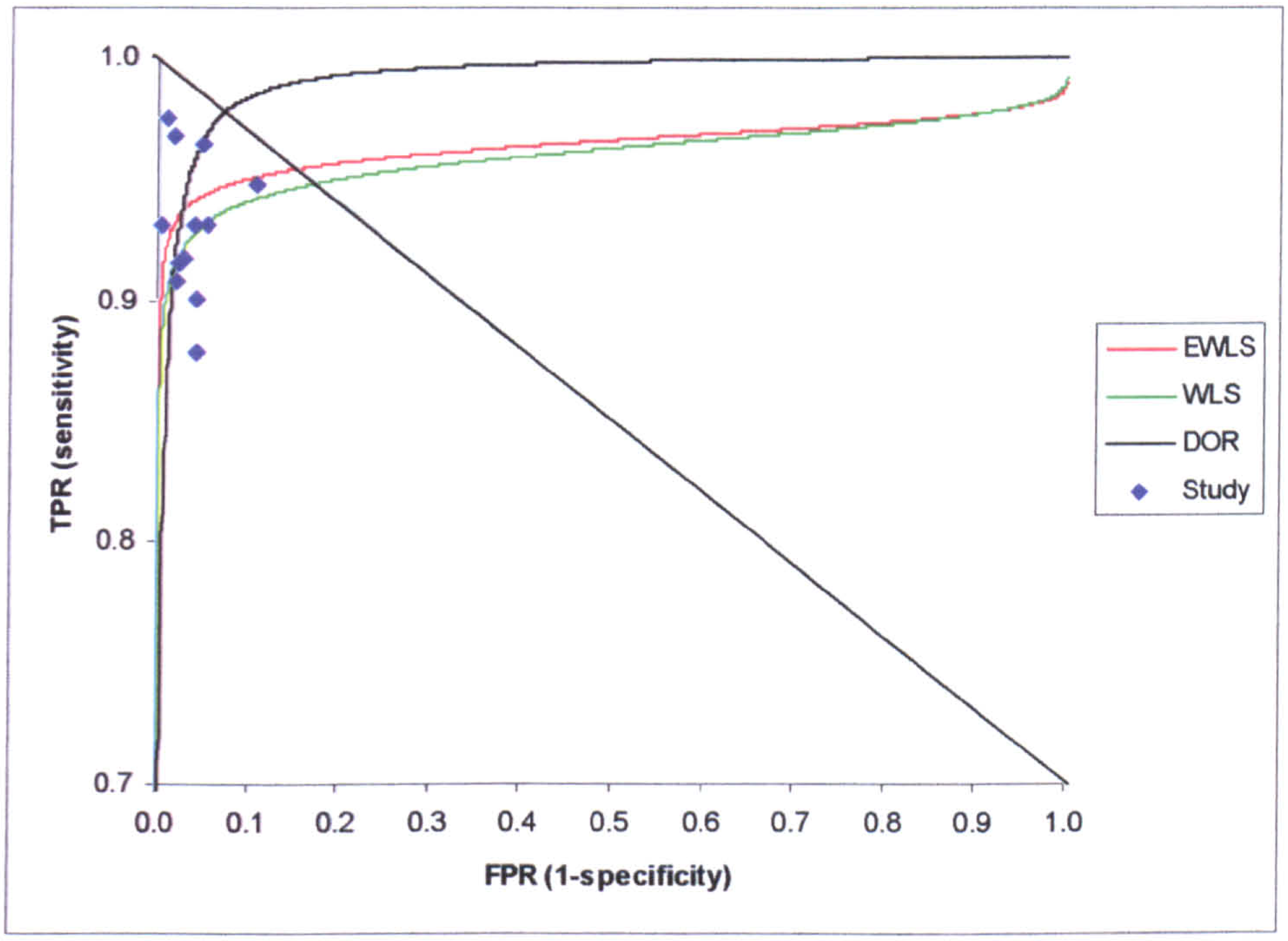
Analyses	α (95% CIs)	P-value	β (95% CIs)	P-value	R^2
EWLS	5.7 (4.7, 6.6)	0.000	-0.7 (-1.4, -0.01)	0.048	0.335
WLS	5.3 (3.8, 6.9)	0.000	-0.6 (-1.9, 0.6)	0.269	0.121

Figure 2.10 D vs S plot



Using the regression coefficient, β, and intercept, α, it was possible to convert back to the conventional axes. Figure 2.11 shows the EWLS and WLS ROC curve and also incorporates the symmetric curve from Figure 2.9. Drawing the "sensitivity = specificity" line demonstrates a departure from symmetry in the EWLS and WLS curves. This is further evidence that the DOR of the individual studies are explained by changes in the diagnostic threshold. Figure 2.11 also shows that sensitivity and specificity are similar for both the symmetric and asymmetric ROC curves in the middle range of the observed studies, but differ at higher sensitivities.

Figure 2.11 Summary ROC curves of radiographer radiograph reporting performance



In conclusion, a summary sensitivity estimate of 92.6 per cent [95 per cent CI: (92.0, 93.2)] and specificity of 97.7 per cent [95 per cent CI: (97.5, 97.9)] suggests that radiographers report plain radiographs to a high level of accuracy. However, the observed heterogeneity provided evidence that alternative meta-analytical techniques should be used to produce a summary estimate as shown in Figure 2.11. Results of the Littenberg & Moses (1993) method showed that the diagnostic threshold explained some of the observed variation but not all of it. Before exploration of other sources of heterogeneity, the secondary objectives of the review are addressed, using summary estimates of sensitivity and specificity. (To avoid repetition, the results of the tests of heterogeneity are not discussed until the concluding section.)

2.9.3.2 Is there a difference in performance between selectively trained radiographers and other health care professionals when reporting radiographs?

Only two studies compared selectively trained radiographers with other health care professionals and both compared radiographers and radiologists reporting A&E radiographs for all body areas. This analysis helped ascertain whether selectively trained radiographers could successfully substitute radiologists reporting A&E radiographs in clinical practice.

Table 2.19 presents the summary sensitivity and specificity estimates for the two professional groups and the findings indicate that the selectively trained radiographers perform slightly better as their pooled point estimates of sensitivity and specificity are higher than those of the radiologists. For both sensitivity and specificity, however, the upper limit of the radiologist 95 per cent CI is similar to the pooled sensitivity and specificity estimates for the radiographers - no evidence to suggest a significant difference in performance. The observed variation in the estimates of sensitivity for radiologists might be explained by Study [5] including registrar and consultant radiologists but Study [4] only including registrar radiologists, resulting in the lower estimate of sensitivity.

In summary, the pooled sensitivity and specificity provided no evidence to claim that radiologists of varying seniority report A&E radiographs for all body areas significantly more accurately than selectively trained radiographers.

Table 2.19 Summary sensitivity and specificity for A&E radiographs

Study	Radiographer		Radiologist	
	Sensitivity (95% CI)	**Specificity (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
[5]	91.7 (86.3, 95.1)	97.3 (95.2, 98.5)	94.9 (90.3, 97.4)	98.0 (96.1, 99.0)
[4]	93.1 (87.0, 96.5)	99.7 (98.9, 99.9)	81.0 (73.0, 87.1)	98.2 (96.9, 99.0)
Total	92.3 (88.5, 94.9)	98.8 (97.9, 99.3)	89.0 (84.8, 92.2)	98.1 (97.1, 98.8)

2.9.3.3 Is there a difference in the performance of selectively trained radiographers reporting radiographs for different types of patient (i.e. non-A&E, A&E)?

Current changes to clinical practice focus on selectively trained radiographers reporting A&E plain radiographs, making it useful to explore their potential for reporting radiographs for a different type of patient, i.e. non-A&E patients. Nine studies were eligible for inclusion in this analysis, although it is important to note that the sample of radiographs included in the two studies used to produce the summary estimates of non-A&E radiograph reporting performance did include some A&E radiographs.

When selectively trained radiographers reported non-A&E radiographs, the summary estimate of sensitivity was 91.9 per cent [95 per cent CI: (90.6, 93.0)] and specificity was 97.4 per cent [95 per cent CI: (97.0, 97.8)]. The summary estimate of sensitivity for A&E radiographs was 92.9 per cent [95 per cent CI: (92.2, 93.6)] and specificity was 97.9 per cent [95 per cent CI: (97.6, 98.1)], indicating that for both sensitivity and specificity, selectively trained radiographers tend to report A&E radiographs more accurately than non-A&E radiographs. The difference in summary estimates between non-A&E and A&E radiographs is 1 per cent or less, so the statistical difference is unlikely to be clinically important.

In summary, there is some evidence to indicate that selectively trained radiographers report A&E radiographs more accurately than non-A&E radiographs though it is unclear whether this difference is either statistically or clinically significant.

2.9.3.4 Is there a difference in the performance of selectively trained radiographers when reporting radiographs for different body areas (i.e. appendicular skeleton, axial skeleton, skeleton, all body areas)?

Evidence indicates that the reporting of radiographs is influenced by body area (Robinson et al, 1999). Variation in radiographer reporting performance for certain body areas may be masked if only overall estimates are presented.

Summary sensitivity and specificity estimates for the different body areas are presented in Table 2.20. The most notable finding suggests that selectively trained radiographers report abnormal radiographs of the appendicular skeleton significantly more accurately than for the axial skeleton, since the 95 per cent upper CI for the pooled axial skeleton sensitivity estimate, 91.5 per cent, does not cross the lower CI for the pooled appendicular skeleton sensitivity estimate, 93.8 per cent. The difference of 4.7 per cent in pooled sensitivity estimates may also be judged clinically important. In contrast, for the specificity estimates, the upper 95 per cent CI for the appendicular skeleton, 98.1 per cent, does not cross the lower CI for the axial skeleton, 98.5 per cent, suggesting that the radiographers reported normal radiographs of the axial skeleton more accurately than for the appendicular skeleton. Nevertheless, there is only a 1.1 per cent difference in the pooled point estimates, which is probably of no clinical significance.

The pooled sensitivity estimates for all body areas, 92.3 per cent, were similar to the pooled estimates for the whole skeleton, 92.6 per cent. So despite evidence by Robinson et al (1999) that chest and abdomen radiographs are more difficult to report, the inclusion of these body areas did not appear to adversely affect radiographers' reporting performance. However, it might be that the other body areas were masking the effect of the chest and/or abdomen radiographs.

Table 2.20 Summary sensitivity and specificity for radiographs of different body areas

Body Area	Sensitivity (95% CI)	Specificity (95% CI)
Appendicular skeleton	94.5 (93.8, 95.2)	97.8 (97.5, 98.1)
Axial skeleton	89.8 (87.8, 91.5)	98.9 (98.5, 99.2)
Skeleton	92.6 (92.0, 93.2)	97.7 (97.5, 97.8)
All body areas	92.3 (88.5, 94.9)	98.8 (97.9, 99.3)

2.9.3.5 Is there a difference in performance between radiographers who have or have not had training when reporting radiographs?

Answering this question was important to establish the effectiveness of training as a method for improving radiographers' performance when reporting radiographs. This should help X-ray Department managers decide whether to invest in such a programme. The only two studies that assessed radiographers without training in a reporting role included A&E radiographs of the skeleton. These were compared with the six studies assessing radiographers with some training reporting A&E radiographs of the skeleton.

Radiographers reporting without training showed a summary estimate of sensitivity of 96.0 per cent [95 per cent CI: (88.9, 98.6)] and specificity was 93.7 per cent [95 per cent CI: (89.3, 96.4)]. For those with some training, the summary estimate of sensitivity was 92.9 per cent [95 per cent CI: (92.2, 93.6)] and specificity was 97.8 per cent [95 per cent CI: (97.6, 98.0)]. The pooled estimate of sensitivity for A&E radiographs of the skeleton was 3.1 per cent higher for the radiographers who had not received any training, but the overlap in CI suggests this difference was not significant. In contrast, the pooled specificity estimate for A&E radiographs of the skeleton was 4.1 per cent higher for the radiographers who had received training. When comparing the CI for the pooled estimates of specificity there is evidence to suggest that training significantly improved radiographers' ability to accurately report normal radiographs. This difference in performance could also be clinically important.

Evidence suggests, then, that radiographers who have received some training can report A&E radiographs of the skeleton significantly more accurately than radiographers who have received none.

2.9.4 Conclusion

Summary estimates of sensitivity and specificity demonstrated that radiographers can report plain radiographs to a high level of accuracy; nor was there evidence that radiologists of varying seniority report A&E plain radiographs for all body areas significantly more accurately than trained radiographers. Furthermore, there was no clear evidence to suggest that a difference in the reporting of A&E and non-A&E radiographs by radiographers with some training is either statistically or clinically significant, though findings indicated that radiographers who have received some training can report A&E radiographs of the skeleton significantly more accurately than radiographers who have not received any training. The meta-analysis, therefore, provides findings to support the substitution of radiologists by trained radiographers in the reporting of A&E plain radiographs and the potential for extending their reporting role to include non-A&E plain radiographs.

However, when first describing the choice of statistical methods used to summarise radiographers reporting performance the need to consider variation between studies in the diagnostic threshold was noted. When using the chi-squared test to detect variation in the results of the twelve studies that addressed the primary objective significant heterogeneity was observed. The subsequent analysis using the Littenberg & Moses (1993) method demonstrated that although this might partly be explained by changes in the diagnostic threshold some variation was left unexplained. When addressing the secondary objectives of the review the results of the chi-squared test again confirmed the presence of significant heterogeneity in estimates of sensitivity and specificity. This was again evidence for not producing single summary estimates of radiographer plain radiograph reporting performance.

In conclusion, before using a combination of qualitative and meta-analytical methods to address the objectives of the review it was first important to explore the sources of heterogeneity as this might further help the interpretation of study results.

2.10 Exploring sources of heterogeneity

If the results of the studies included in the meta-analysis about radiographers' plain radiograph reporting performance were homogenous, a summary estimate should provide reliable evidence. However, the observed heterogeneity questioned the validity of drawing conclusions from combining studies. Because the observed heterogeneity was only partly as a result of the threshold effect (as discussed in section 2.9.3.1) there follows a discussion of the statistical method used to further investigate sources of heterogeneity.

2.10.1 Method of investigating sources of heterogeneity

If assumption can not be made, as found in section 2.9, that the summary ROC curves are symmetrical, sources of heterogeneity can be investigated by extending the Littenberg & Moses (1993) regression method. Adding a covariate to the equation for each potential effect modifier can provide evidence of whether variability in accuracy is explained by that covariate, as described by Deeks (2001 p.277). For example, in the equation below, *D* is the natural logarithm of the DOR and is the dependent variable for a single study. The intercept is α and is interpreted as the common DOR and the independent variable representing the diagnostic threshold in a study is *S*. The regression coefficient, β_0, expresses variation in *D* across studies and provides an estimate of the extent to which *D* is independent of the threshold. For example, adding a binary covariate to indicate studies that used a valid reference standard (*VRS*), where *VRS* is 1 in studies that did use a *VRS* and 0 when otherwise, allows a hypothesis of equal DOR to be tested. The estimate of β_1 indicates the performance of radiographers in studies that *did* use a *VRS*, relative to the performance of radiographers when studies lacked a *VRS*. The resulting parameter estimates of the covariates can be interpreted after antilogarithm transformation as relative DOR (RDOR). If the RDOR ($\exp(\beta_1)$) is smaller than one, it indicates that studies with *VRS* yield smaller estimates of accuracy than studies which do not.

$$D = \alpha + \beta_0 S + \beta_1 VRS$$

2.10.2 *Application of Littenberg & Moses extended regression method*

Variation in the results of primary studies of diagnostic tests may be caused by chance, but can also reflect true heterogeneity (Lijmer et al, 2002). Differences in study populations, setting, or cutpoint for abnormality, or any combination of such factors (Thompson, 1994), are possible clinical sources of heterogeneity. In addition, heterogeneity in results can be caused by flawed designs in some studies: methodological or artefactual heterogeneity (Irwig et al, 1995). The covariates chosen *a priori* for this analysis include the biases, methodological standards, and other clinical characteristics of the studies referred to in the data extraction form (see Annex 2.2).

Table A2.5.1 in Annex 2.5 lists these covariates and whether they were absent (A), present (C), unclear (B), or not applicable (N/A). Not all covariates were included in the subsequent analysis, as Table A2.5.1 shows some did not vary between studies. For example, intra-arbiter (E6) and inter-arbiter variability (E7) were present in all twelve studies, whereas popularity bias (B2) was absent in all studies. Some covariates were excluded because they were not applicable to any study (e.g. C2, E1). Table A2.5.2 lists the covariates that were included in the regression, and their values. When first conducting the regression the covariates were classified as binary variables, so the criterion scored as a categorical variable, e.g. the validity of the arbiter (H5) was scored as, invalid arbiter = 0, valid arbiter = 1. Table A2.5.3 presents the data included in the regression analyses. When it was not clear (B) whether a bias was absent (A) it was assumed present (C). Analyses were run using SPSS® (Version 10.0 SPSS Inc. 2001. Chicago, Illinois, USA).

Both EWLS and WLS analyses were performed, the choice of which did not influence whether a covariate was important for explaining variation in *D*. Table A2.5.4 presents the findings when weighting was used. In performing n tests, it is recommended that the significance level is divided by n to arrive at the uncorrected probability to determine statistical significance (i.e. 0.05/n) (Gore, 1982). In this case, a value of $P<0.002$ would be required (i.e. 0.05/30). This is sometimes referred to as the Bonferroni correction and was used as a cautious approach to interpreting results of multiple significance testing. Table A2.5.4 shows only one variable met this criterion (G1), which is the inclusion of a sample size calculation (P=0.001).

Results of the weighted regression linear analyses are given in Table 2.21, including and excluding G1 as a covariate in the equation. R^2 is 0.121 when G1 is not included in the model but 0.756 when G1 is included, suggesting that a large proportion of the variation in the dependent variable *D* is explained by the presence of G1. Furthermore, the resulting RDOR is 14.01 indicating that the study using a sample size calculation yielded a higher estimate of accuracy than the other studies. Table 2.22 shows similar findings when unweighted regression analysis was performed.

Table 2.21 Weighted summary ROC models with and without G1

Variable	Coefficient (95% CI)	
	Basic Model	Model Including G1
Intercept	5.33 (3.79, 6.87)	5.23 (4.36, 6.10)
S	-0.64 (-1.86, 0.58)	-0.56 (-1.24, 0.13)
G1	not included	2.64 (1.41, 3.87)
R^2	0.121	0.756

Table 2.22 Unweighted summary ROC models with and without G1

Variable	Coefficient (95% CI)	
	Basic Model	Model Including G1
Intercept	5.67 (4.74, 6.60)	5.54 (4.78, 6.31)
S	-0.70 (-1.38, -0.01)	-0.63 (-1.20, -0.07)
G1	not included	2.23 (0.26, 4.20)
R^2	0.335	0.615

Only one Study [28] used a sample size calculation (G1). This was a multi-centre study with ten selectively trained radiographers reporting A&E skeletal plain radiographs across four NHS Trusts. Results of studies that did not use a sample size calculation were pooled to test whether the exclusion of Study [28] helped to explain the evident heterogeneity when all twelve studies in the meta-analyses were combined. Table 2.23 shows that for both sensitivity and specificity the 95 per cent CI do not overlap, but the results of combining the studies when G1 was absent were still significantly heterogeneous for both sensitivity (X^2 = 35.4, df = 10, P<0.001) and specificity (X^2 = 97.2, df = 10, P<0.001). It is important to note that the significance of G1 for explaining variation in *D* could be a spurious finding, because, by chance, the only study that used a sample size calculation produced one of the highest estimates of sensitivity and specificity.

Table 2.23 Sensitivity and Specificity by presence of G1

Variable	G1 (present)	G1 (absent)
Sensitivity	97.47 (96.43, 98.21)	91.68 (90.98, 92.34)
Specificity	99.27 (99.02, 99.46)	97.15 (96.89, 97.40)

The preceding analyses treated the additional covariates as binary variables even though, as shown in Table A2.5.1, many of these covariates originally had more than two categories (e.g. H4, H5). A simple conclusion could then be made about whether the presence or absence of a covariate explains variation in *D*. The analysis was repeated for both unweighted and weighted models, but without the conversion of all covariates into binary variables.

For this analysis, G1 did not change as it already was a binary variable. F2 was the only other variable that now produced a significant finding, where P<0.01. This represented reference standard review bias, which is present when the reference standard is not blind to the report of the person under evaluation. Table 2.24 shows the mean values for *D* depending on whether the bias was absent, unclear, or present using weighted analyses. The R^2 value was 0.903 for this model and the difference in *D* between the three subgroups was highly significant (P=0.000). Table 2.25 shows that all the pairwise comparisons between the three subgroups resulted in a significant difference in mean values for *D*.

Table 2.24 Descriptive statistics for F2

F2	Mean	Standard Deviation	N
Absent	6.01	1.78	8
Unclear	5.10	.	1
Present	8.37	2.08	3
Total	6.08	4.05	12

Table 2.25 Pairwise comparisons for F2

(I) F2	(J) F2	Mean Difference (I-J)	Std. Error	Sig.	95% CI for Difference	
					Lower Bound	Upper Bound
Absent	Unclear	.903	.265	.009	.291	1.515
	Present	-2.282	.338	.000	-3.062	-1.503
Unclear	Absent	-.903	.265	.009	-1.515	-.291
	Present	-3.185	.400	.000	-4.109	-2.262
Present	Absent	2.282	.338	.000	1.503	3.062
	Unclear	3.185	.400	.000	2.262	4.109

The analysis was then conducted including both G1 and F2 as additional covariates. Table 2.26 shows that most of the variation in *D* was explained by reference standard review bias (F2). *S* as the measure of diagnostic threshold also made a statistically significant contribution to explaining variation in *D*, but G1 did not make a significant contribution to this model. There was not enough data to investigate an interaction. The R^2 value was 0.932, indicating that the variables included in the model explain most of the variation in the dependent variable *D*.

Table 2.26 Weighted model including G1 and F2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	168.368	4	42.092	24.141	.000	.932
Intercept	938.725	1	938.725	538.375	.000	.987
S	14.237	1	14.237	8.165	.024	.538
G1	5.293	1	5.293	3.036	.125	.302
F2	31.777	2	15.889	9.112	.011	.722
G1 * F2	.000	0	.	.	.	.000
Error	12.205	7	1.744			
Total	8951.503	12				
Corrected Total	180.574	11				

Findings from this analysis suggest that when the covariates were not converted into binary variables, F2 explained most of the variation, although when F2 was classified as a binary variable and regressed on *D*, there was no indication that it significantly explained variation in *D* (P=0.694). Furthermore, it was only for Study [16] that F2 could not be categorised as absent or present. Study [16] had the lowest estimate of sensitivity over all, so as with G1, the result is likely to be spurious, as a change in categorising one study brought about a significant finding.

2.10.3 Conclusion

With only twelve studies included in this analysis, the results must be interpreted with caution as to the importance of sample size calculations and reference standard review bias explaining the sources of variation. Perhaps the most notable finding was the extent to which variation in the diagnostic threshold, or *S*, between studies explains variation in *D*. It seems plausible that the threshold for defining normal or abnormal varies between studies when there are different professionals using different reporting styles, different reference standards of varying validity, and different arbiters using different criteria to judge concordance between reports. Human heterogeneity inherent in the process of reporting radiographs and comparing reports for concordance might explain variation in the diagnostic threshold between studies. The next section will investigate one more bias before interpreting the results of this review.

2.11 Investigating publication bias

In this review, several methods were employed to identify studies, including searching electronic databases, handsearching journals and conference abstracts, personal communication and reading reference lists. Searches of the SIGLE database, handsearches of the BJR annual congress supplements and personal communication were methods of locating grey literature. Twenty-four of the 35 studies in the review were defined as grey literature (69 per cent), as were nine of the twelve studies (75 per cent) included in the meta-analysis. This high percentage of grey literature suggests a comprehensive search strategy was employed to identify all eligible studies. Nevertheless, it was important to consider the presence of publication bias, as this can exaggerate estimates of effectiveness (McAuley et al, 2000).

Publication bias has been defined as a bias introduced by the nature of publication depending on the direction of results (Egger et al, 2001), and is more likely to occur in smaller studies, which tend to be of lower methodological quality (Sterne et al, 2000). Grey literature studies are unpublished, have limited distribution and/or are not included in bibliographical retrieval systems (Last, 1995). To visually investigate the possible presence of publication bias a funnel plot can be produced. This is a simple scatter plot of the estimate of effect in individual studies (horizontal axis) against some measure of precision such as study size (vertical axis) (Sterne et al, 2001). The plot is termed 'funnel' as due to random variation, smaller size studies will have a wider distribution of results than larger studies. Due to the nature of the publication, smaller negative studies may not appear in the literature, possibly leading to asymmetry in the funnel and suggesting the possibility of publication bias.

To investigate the presence of publication bias in the studies included in the meta-analysis, plots were produced to present sensitivity and specificity respectively against precision (1/standard error). To explore the possibility of a ceiling effect the plots were also produced so the scale of the x-axis was the cubed of sensitivity and specificity. Figure 2.12 shows a funnel shape with smaller studies demonstrating wider variation in sensitivity than the larger studies, suggesting that publication bias is not present. Changing the x-axis as shown in Figure 2.13 helps to disperse the plots but does not affect interpretation.

Figure 2.12 Funnel plot for sensitivity

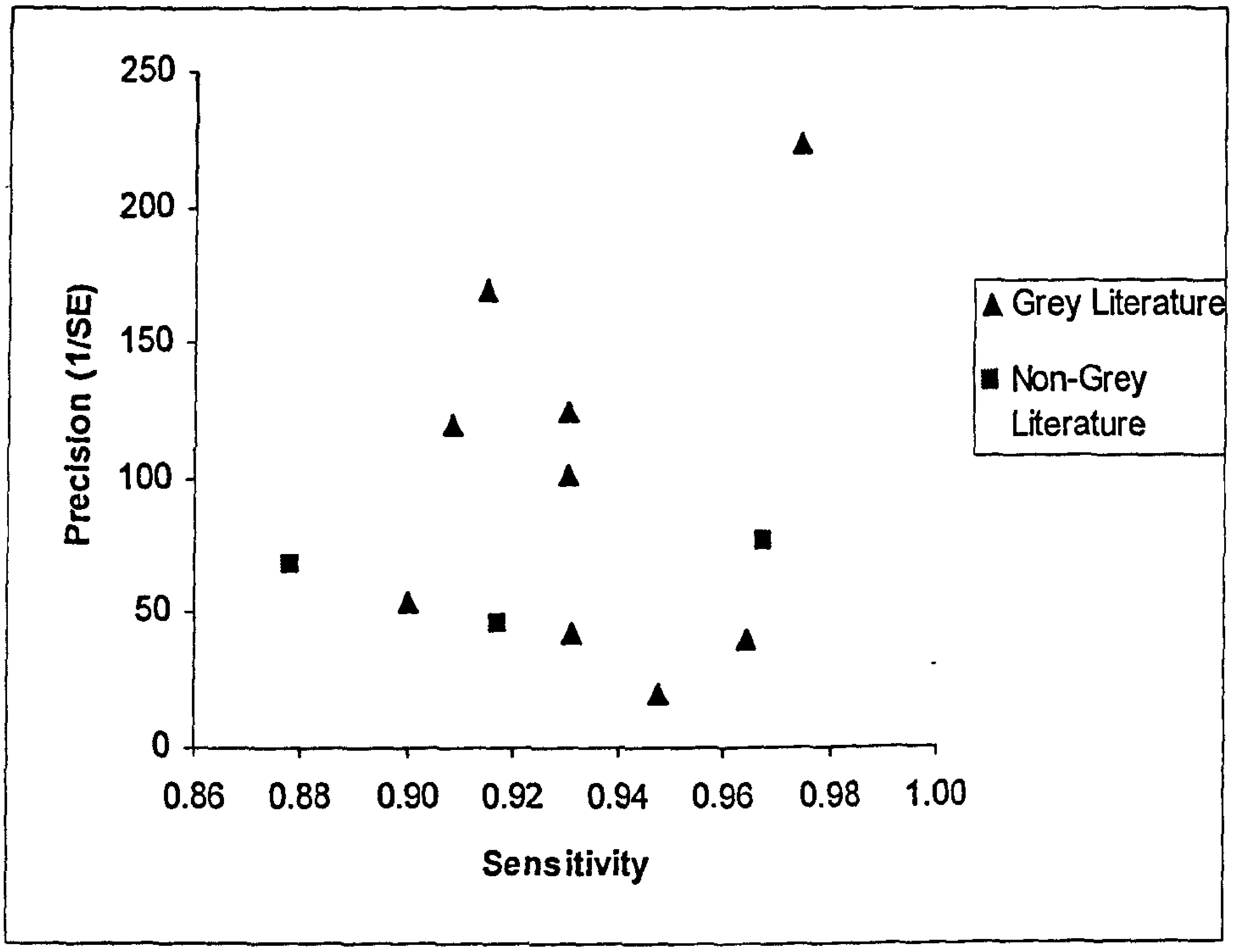


Figure 2.13 Funnel plot for sensitivity cubed

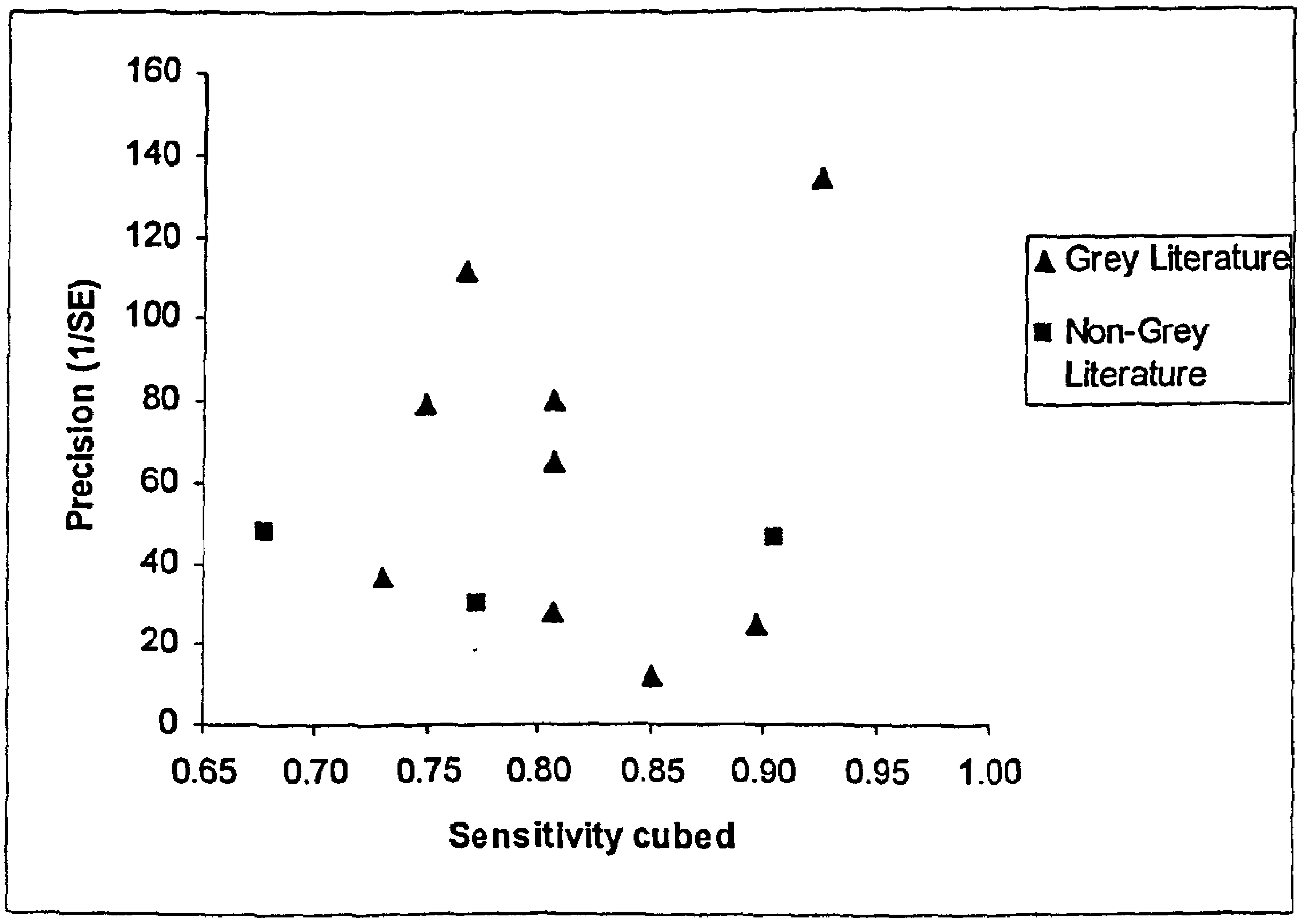


Figure 2.14 also seems to show a funnel shape with the smaller studies demonstrating wider variation in specificity although the variation is not as well distributed as for sensitivity. For the smaller size studies, there is also a cluster of grey literature which indicates the possible presence of publication bias, as these studies have lower estimates of specificity than those published. There are also large grey literature studies with high sensitivity and specificity which are not published. Moreover, the largest of all studies was grey literature and had the highest sensitivity and one of the highest specificities. It would seem that the direction of results does not affect the publication or non-publication of findings. Again, changing the x-axis as shown in Figure 2.15 helps to disperse the plots but does not affect interpretation.

Figure 2.14 Funnel plot for specificity

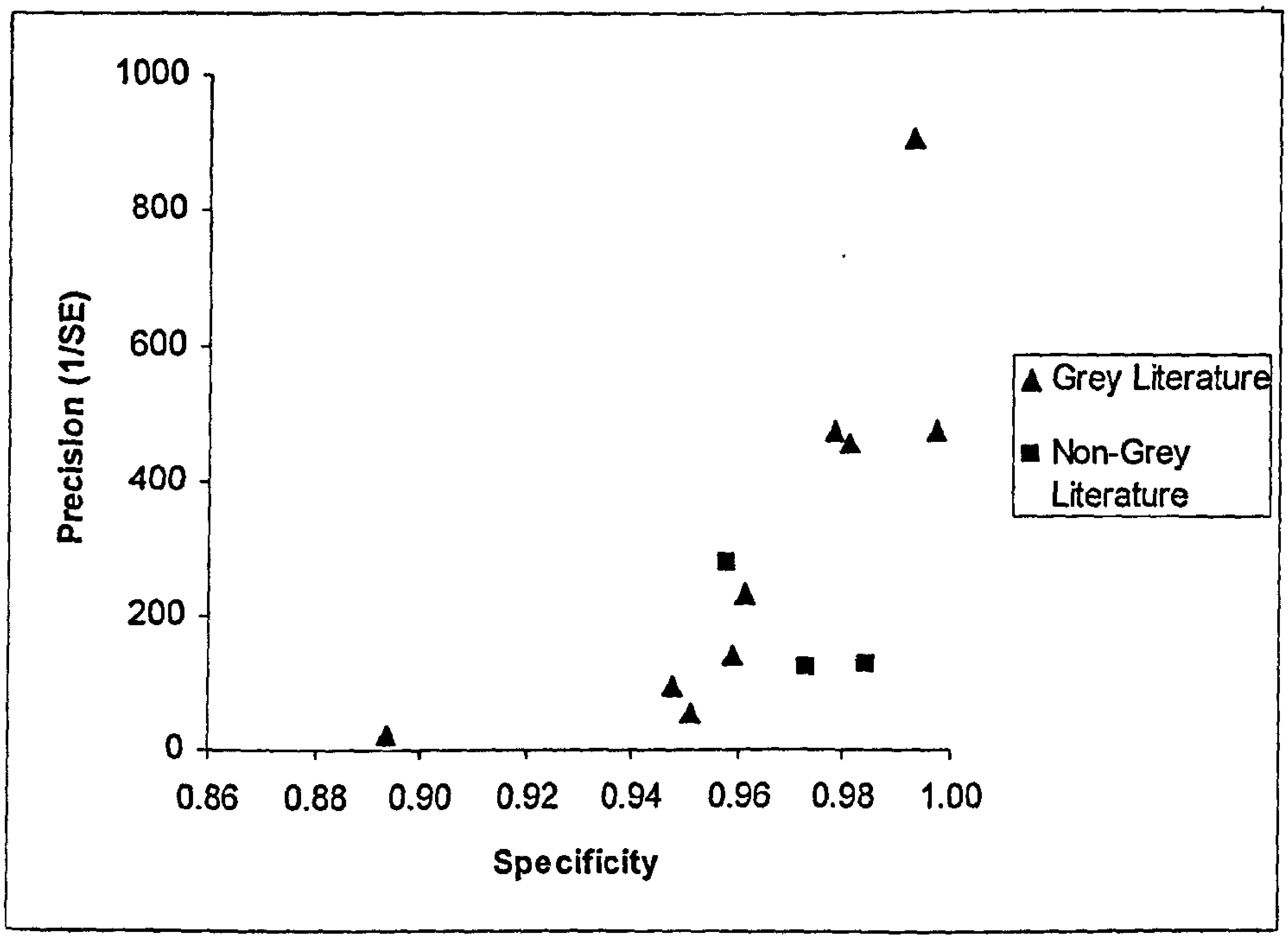
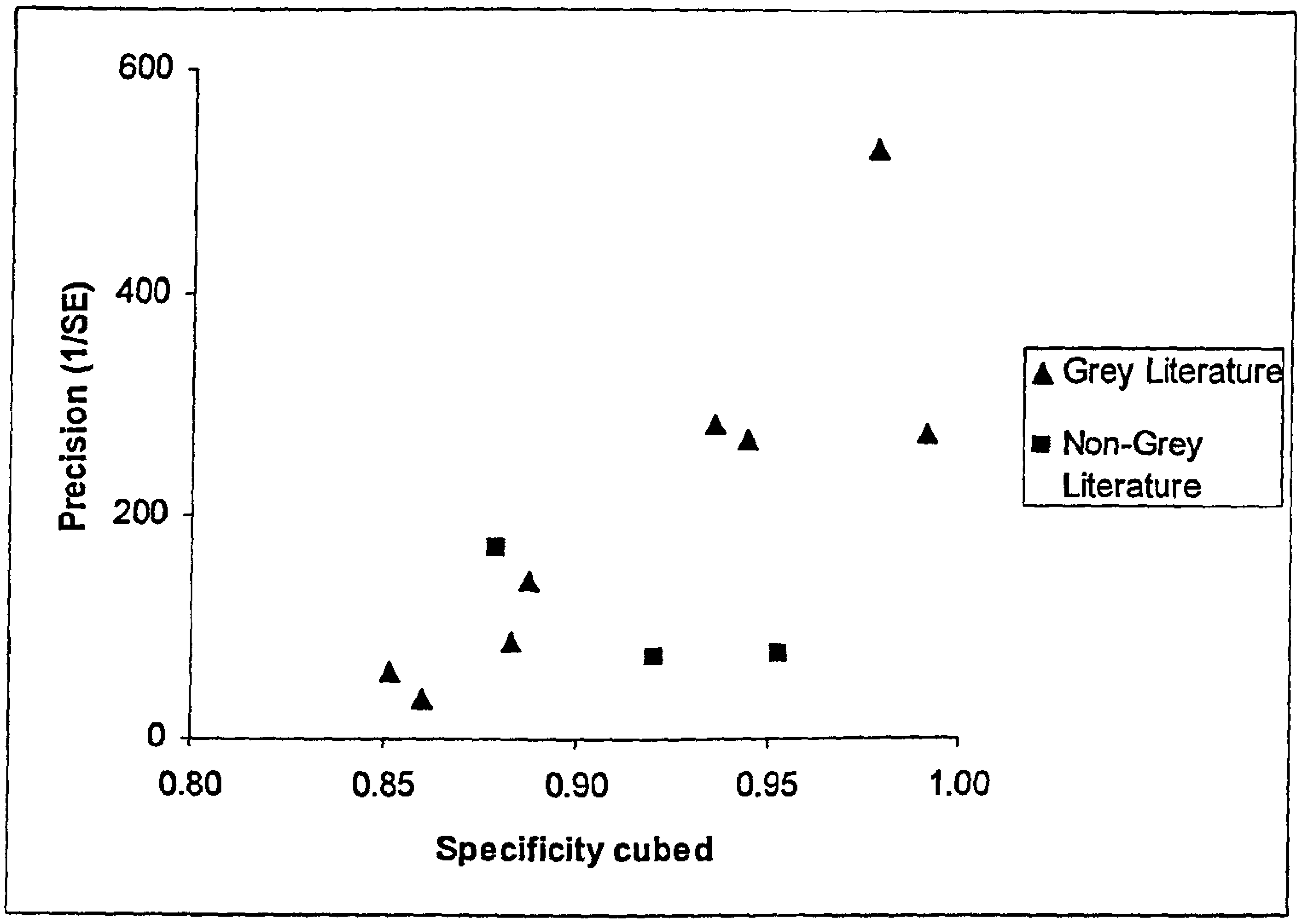


Figure 2.15 Funnel plot for specificity cubed



Visual displays can only be interpreted subjectively. Further support for the suggestion that there is no publication bias was sought using the extended Littenberg & Moses (1993) method. It found there was insufficient evidence to sustain the hypothesis that the covariate 'Grey Literature' explains variation in the accuracy with which radiographers report plain radiographs (P=0.195). The same findings emerged when unweighted analysis was performed (P=0.751). Grey literature should cause variation in *D* if the results of the studies included in the meta-analysis were influenced by whether or not they were grey literature. Both graphical and statistical evidence suggests that publication bias should not be an important factor for confounding the interpretation of results from this review. One reason why so many studies are not published in a peer-reviewed journal is that they are often the results of audits, performed after radiographer's receive training to quality assure their performance in clinical practice, and are not intended for publication. Also, several of the studies were identified as abstracts to conferences and not subsequently developed into a paper.

2.12 Interpretation of results

Results from the previous sections about the qualitative overview (section 2.8), quantitative synthesis (section 2.9), sources of heterogeneity (section 2.10) and publication bias (section 2.11) will be summarised here to address all the objectives of the review. In the following discussion, the letters in square brackets refer to the quality of evidence and the numbers in superscript refer to the studies providing this evidence.

2.12.1 Assessing the plain radiograph reporting performance of radiographers

Ten of the eleven diagnostic accuracy studies assessed radiographers in a reporting role and seven of the ten used marking schemes to calculate sensitivity and specificity as well as accuracy. Five studies were of high quality [A] [18, 17, 11, 22, 10] and two of moderate quality [B] [13, 14]. The radiographers' plain radiograph reporting accuracy ranged respectively from 62 per cent to 95 per cent and from 65 per cent to 95 per cent. Such a broad range of reporting accuracy seems to be explained by variation in the type of patient, body area and training received and is discussed further when addressing the secondary objectives of the review.

Eleven of the fifteen diagnostic performance studies also assessed radiographers in a reporting role. Nine were of moderate quality [B] [28, 30, 12, 26a, 15, 19, 9, 8, 32] and two were of low quality [C] [21, 34] and the radiographers' reporting accuracy ranged from 94 per cent to 99 per cent. Five of the nine diagnostic outcome studies assessed radiographers in a reporting role and were of varying quality [A] [26b], [B] [5, 4, 31] [C] [16]. For these studies the radiographers' plain radiograph reporting accuracy ranged from 91 per cent to 99 per cent.

Although the observed heterogeneity questions the validity of pooling data from studies conducted during clinical practice, the summary estimate of sensitivity and specificity were found respectively to be 92.6 per cent [95 per cent CI: (92.0, 93.2)] and 97.7 per cent [95 per cent CI: (97.5, 97.9)]. These findings indicated that radiographers report plain radiographs during clinical practice at a higher level of accuracy and more consistently than when reporting in controlled conditions.

2.12.2 *Radiograph reporting performance of selectively trained radiographers compared with other health care professionals*

One of the review's secondary objectives was comparison of selectively trained radiographers' performance with other professional groups, in an attempt to inform which professions should be performing this task. Two diagnostic accuracy studies conducted in controlled conditions found radiographers with training could report A&E radiographs of the skeleton significantly better (P<0.02), or with little significant difference from radiologist registrars [B] [23, 27] and for one study, significantly more accurately than A&E nurses (P<0.05) [A] [10]. Another diagnostic accuracy study found that radiographers with some training reported A&E radiographs for various body areas more accurately than radiographers without training and junior radiologists [A] [22].

One diagnostic outcome study provided evidence that selectively trained radiographers can report A&E radiographs for all body areas with no significant difference (P>0.2) from radiologists of varying seniority [B] [5]. In contrast, another study found that selectively trained radiographers reported A&E radiographs for all body areas significantly more accurately than radiologist registrars (P<0.001) [B] [4]. Two other studies of varying quality demonstrated that trained radiographers report A&E plain radiographs of the skeleton to a similar degree of accuracy as consultant radiologists [A] [26b] [C] [16], though even the high quality study was undermined by uncertainty about the presence of some bias concerning the application of the arbiter (F7, F8, F11).

In summary, the diagnostic accuracy studies provided evidence that selectively trained radiographers can report A&E radiographs significantly more accurately than A&E nurses with or without training, and than other radiographers without training [A] [10, 22]. Furthermore, radiographers with training can report A&E radiographs similar to radiologists of varying seniority [B] [23, 27], though all these studies were performed under controlled conditions with a highly selective choice of radiographs and could be described as only assessing the efficacy of radiographer reporting. It was essential that the diagnostic outcome studies performed during clinical practice, which provided evidence on the effectiveness of radiographer reporting, underpinned the diagnostic accuracy study findings. Notably, four studies of varying quality provided evidence that selectively trained radiographers reported A&E radiographs to a comparable level of accuracy as radiologists of varying seniority [A] [26b] [B] [5, 4] [C] [16]. Despite the evident heterogeneity between the two studies that were eligible for inclusion in the meta-analysis, the pooled estimates of sensitivity and specificity did not provide evidence that radiologists of varying seniority report A&E radiographs for all body areas significantly more accurately than radiographers [B] [5, 4].

2.12.3 Radiograph reporting performance of selectively trained radiographers for different categories of patients (e.g. A&E, non-A&E)

Another secondary objective was to explore whether selectively trained radiographers performance was similar for different categories of patients. Evidence from four diagnostic accuracy studies of high quality indicated that selectively trained radiographers reported A&E plain radiographs of the skeleton between 93 per cent and 95 per cent accuracy [A] [18, 17, 22, 10]. These are the four highest quality studies included in the review. There were also three studies of moderate quality that provided evidence of selectively trained radiographers reporting A&E radiographs of the skeleton [B] [13, 23, 27], but the latter two studies did not use conventional scoring systems and the first study shows that the radiographers only report at 81per cent accuracy. The sample of radiographs used in this study as part of a training assessment exercise might explain the result. Two of the high quality studies also provided evidence that selectively trained radiographers reported non-A&E radiographs of the skeleton between 93 per cent and 94 per cent accuracy [A] [18, 17]. Another showed that selectively trained radiographers reported a general sample of radiographs from a variety of sources at 95 per cent accuracy. So the evidence from high quality studies, but performed in controlled conditions, suggests that radiographers who have received some training report A&E and non-A&E radiographs of the skeleton to a similar level of accuracy.

Seven diagnostic performance studies, of moderate to low quality, provided evidence that selectively trained radiographers report A&E plain radiographs of the skeleton between 94 per cent and 99 per cent accuracy [B] [28, 30, 12, 15, 19] [C] [21, 34] and two diagnostic outcome studies of varying quality show that radiographers with some training report A&E plain radiographs of the skeleton at 97 per cent accuracy [A] [26b] [C] [16]. Two other diagnostic outcome studies also show that radiographers with some training report A&E plain radiographs for all body areas between 96 per cent and 99 per cent accuracy [B] [5, 4]. Three diagnostic performance studies showed radiographers with some training reported radiographs of the skeleton from various sources including A&E between 96 per cent and 98 per cent accuracy [B] [26a, 9, 8]. There also appears to be evidence, but from studies performed during clinical practice, that radiographers report plain radiographs from different sources to a similar level of accuracy.

The findings from the meta-analysis, despite the evident heterogeneity, were that the pooled estimates of sensitivity and specificity for non-A&E radiographs were respectively 91.9 per cent [95 per cent CI: (90.6, 93.0)] and 97.4 per cent [95 per cent CI: (97.0, 97.8)] [B] [9, 8] and the pooled estimates for non-A&E radiographs were respectively 92.9 per cent [95 per cent CI: (92.2, 93.6)] and 97.9 per cent [95 per cent CI: (97.6, 98.1)]. This would indicate that selectively trained radiographers report A&E radiographs more accurately than non-A&E radiographs, although the difference in summary estimates of accuracy between non-A&E and A&E radiographs is 1 per cent or less, making any statistical difference unlikely to be clinically important.

2.12.4 Radiograph reporting performance of selectively trained radiographers for different body areas

That radiographers with some training can report radiographs of the skeleton between 93 per cent and 95 per cent accuracy [A] [18, 17, 22, 10] was evident from four diagnostic accuracy studies. One of these also demonstrated that radiographers with some training report A&E radiographs of the chest and abdomen respectively at 83 per cent and 78 per cent accuracy [A] [22]. In the same study, junior radiologists reported the chest and abdomen respectively at 73 per cent and 58 per cent accuracy. Three moderate quality studies provided evidence of radiographers with some training reporting A&E radiographs of the skeleton [B] [13, 23, 27], but the latter two studies did not use conventional scoring systems and the first showed that the radiographers only reported at 81 per cent accuracy. Overall, evidence from studies conducted in controlled conditions was that both selectively trained radiographers and radiologists found it more difficult to report chest and abdomen A&E radiographs compared with the skeleton.

Of those studies conducted during clinical practice, four diagnostic performance studies showed that radiographers with some training reported radiographs of the appendicular skeleton from 94 per cent to 98 per cent accuracy [B] [30, 26a, 8] [C] [21]. Six diagnostic performance studies provided evidence that radiographers with some training reported radiographs of the whole skeleton between 95 per cent and 99 per cent accuracy [B] [28, 12, 15, 19, 9] [C] [34].

One diagnostic outcome study also showed that radiographers with training reported A&E radiographs of the appendicular skeleton at 97 per cent accuracy [A] [26b] and another for the whole of the skeleton at 97 per cent accuracy [A] [16]. In these studies, consultant radiologists reported these radiographs slightly more accurately at 98 per cent [A] [26b] and 99 per cent [A] [16]. Finally, two other diagnostic outcome studies showed that radiographers with training reported A&E plain radiographs for all body areas between 96 per cent and 99 per cent accuracy [B] [5, 4] and the same studies showed that radiologists of varying seniority reported these radiographs between 96 per cent and 97 per cent accuracy. So results from studies performed during clinical practice support the findings of those conducted in controlled conditions, that radiographers with training report radiographs of the skeleton at a high level of accuracy.

Interestingly, the results of two of the diagnostic outcome studies showed that selectively trained radiographers reported A&E radiographs for all body areas (including chest and abdomen) between 96 per cent and 99 per cent accuracy [B] [5, 4] - a similar level to the studies that assessed radiographer performance when reporting radiographs of the skeleton. Indeed, whilst noting the evident heterogeneity, the meta-analysis showed that the pooled estimates of sensitivity for all body areas at 92.3 per cent [95 per cent CI: (88.5, 94.9)] were similar to the pooled estimates of sensitivity for the skeleton, 92.6 per cent [95 per cent CI: (92.0, 93.2)]. A possible explanation is because both diagnostic outcome studies were prone to biases concerning the application of the reference standard (D1, D2) and the arbiter (F7, F8) [B] [5, 4] resulting in elevated estimates of radiographer performance. Nevertheless, the findings indicate that radiographers with some training reported radiographs for all body areas to a high level of accuracy. The limited evidence available also indicated that radiographers with training reported A&E plain radiographs of the chest and abdomen less accurately than for the skeleton, but the same also applies to radiologists.

2.12.5 *The effectiveness of training programmes for improving radiographer radiograph reporting performance*

The final secondary objective was to assess the effectiveness of training programmes as a method for improving radiographer radiograph reporting. Two of the diagnostic accuracy studies provided evidence that radiographer reporting accuracy of A&E plain radiographs of the skeleton improved with training [A] [17] [B] [13]. Two others provided high quality evidence that radiographers with training reported A&E plain radiographs of the skeleton more accurately than radiographers without training [A] [22, 10] and one of these showed the same for the reporting of A&E radiographs of the chest and abdomen [A] [22]. One diagnostic outcome study, although of low quality, also provided evidence that training programmes can significantly improve radiographer's accuracy when reporting A&E radiographs of the skeleton ($P<0.001$) [C] [16]. Most of these studies were conducted in controlled conditions, but nevertheless provided consistent evidence that training radiographers to report A&E radiographs improved accuracy.

Results from the meta-analysis showed the summary estimates of sensitivity and specificity for radiographers with no training were respectively 96.0 per cent [95 per cent CI: (88.9, 98.6)] and 93.7 per cent [95 per cent CI: (89.3, 96.4)]. Those for radiographers with training were respectively 92.9 per cent [95 per cent CI: (92.2, 93.6)] and 97.8 per cent [95 per cent CI: (97.6, 98.0)]. The summary estimate of sensitivity for radiographers without training was 3.1 per cent higher than for the radiographers who had received some training, neither was there any overlap in CI for the summary estimates, suggesting this was a significant difference. In contrast, the summary estimate of specificity for radiographers with training was 4.1 per cent higher than for radiographers without training. The absence of overlap in CI for the summary estimates suggests this was a significant difference and further evidence to suggest that radiographers who have received some training can report A&E radiographs of the skeleton significantly more accurately than radiographers who have not received any training. However, the evident heterogeneity limits the validity of this conclusion.

2.12.6 Plain radiograph reading performance of radiographers in a red dot or triage role

One of the supplementary objectives of the review was to assess radiographer reading performance other than in a reporting role. The only diagnostic accuracy study that assessed radiographers in a different role found that training significantly improved their ability to accurately triage radiographs of the chest (including thoracic cage) for A&E, in-patients and out-patients [C] [2]. This study was one of the lowest quality of all those included in the review, and the absence of a control group and selection of radiographs before and after training seriously undermined its validity.

Of the fifteen diagnostic performance studies, two low quality studies indicated that radiographers can red dot A&E radiographs for all body areas at 90 per cent accuracy or the skeleton only at 92 per cent accuracy [C] [25, 33]. One also demonstrated that following training, radiographer accuracy improved from 89 per cent to 92 per cent when red-dotting A&E radiographs of the skeleton [C] [33]. One diagnostic performance study also found that radiographers correctly red-dotted 98 per cent of radiographs of the chest (including thoracic cage) for in-patients at a cardio thoracic centre [B] [29]. Another study found that radiographers accurately triaged 92 per cent of A&E radiographs of the skeleton - significantly more accurate than their triaging of chest/abdomen radiographs from the same source ($P<0.001$) [B] [7].

Diagnostic outcome studies showed that radiographers could red dot or triage A&E radiographs for all body areas between 91 per cent and 94 per cent accuracy [B] [20, 6] and skeletal radiographs only between 90 per cent and 96 per cent [C] [24, 1]. Two of these also provided evidence that radiographers without training could red dot or triage A&E radiographs for all body areas as accurately or more accurately than casualty officers [B] [20, 6]. The other two studies found that the combined performance of radiographers and another professional group, such as senior house officers or casualty nurses, improved accuracy when reading A&E plain radiographs of the skeleton [C] [24, 1]. Whether this was significantly different is unknown.

In summary, there were nine studies of varying quality that demonstrated radiographers accurately red dotted or triaged plain radiographs [B] [29, 7, 20, 6] [C] [2, 25, 33, 24, 1], but it should be noted that four of the studies were the lowest ranking quality studies included in the review. In particular, the two studies that provided evidence about the combined performance of radiographers with other health care professionals are the two of lowest quality [C] [25, 33]. Even the highest quality studies assessing radiographers' red-dotting radiographs of the chest for in-patients at a cardio-thoracic centre was prone to several sources of bias [B] [29], including reference standard review bias (F2) and biases concerning the application of the arbiter (F6, F7, F8). In addition, the criteria for a valid reference standard (H4) and arbiter (H5) were not met.

2.12.7 Threats to the validity of plain radiograph reading performance studies

Identification of threats to the validity of studies that assess radiographer plain radiograph reading performance was a further supplementary objective. Its rationale was to help assess the validity of studies included in the review for the purpose of data synthesis and to inform the design and conduct of the primary studies presented in Chapters 4 to 6 and resulted with the development of the data extraction form presented in Annex 2.2. The frequency with which these threats were present in terms of bias and methodological standards is discussed here, with raw data presented in Annex 2.6. The reviewers have recently published papers that discuss this in more detail, but only studies before May 1999 were included in these papers as the search strategy was updated more recently to the end of October 2002 (Brealey et al, 2002a, b).

In summary, across all studies the frequency of bias in radiograph selection (0 per cent to 38 per cent), observer selection (0 per cent to 8 per cent), and application of the reference standard (17 per cent to 20 per cent) was quite low, though most studies failed to assess intra- and inter-observer/arbiter variability (93 per cent to 100 per cent). Many biases were also present concerning independence in decision-making, including: the arbiter being one of the observers or reference standard (64 per cent); the reference standard reporting radiographs with knowledge of the observers report (69 per cent); the arbiter using radiographs when judging reports for concordance (83 per cent); and the arbiter being aware of which report was made by whom (93 per cent to 97 per cent). These biases are important since there is evidence and consensus that access to certain knowledge can adversely affect decision-making (Sackett et al, 1991). The concept of independence in decision-making is analogous to blinding in randomised trials when it can be essential that clinicians, patients and statisticians are unaware of treatment assignment (Moher et al, 2001).

The percentage of studies that met the methodological standards are as follows:

• performed a sample size calculation	6%
• defined what was a normal and abnormal report	89%
• described the sequence of events through which films passed before reporting	92%
• analysed individual groups of observers within a combination of groups	50%
• appropriate choice of reference standard	77%
• appropriate choice of arbiter	57%
• appropriate use of a control	36%
• analysis of pertinent clinical subgroups (e.g. body areas patient type)	47%
• availability of data for re-calculation	64%
• presentation of indeterminate results	71%

These findings indicated wide variation both in the presence of bias and adherence to methodological standards in studies of radiographers' radiograph reading performance. The regression analyses presented in section 2.10.2 showed that the presence or absence of sample size calculations (G1) and reference standard review bias (F2) significantly affected study findings, though the small number of studies in the sample probably produced these findings by chance. Careful consideration of all the biases and standards identified is an essential component of study quality and the validity of the evidence-base used to underpin radiographer reporting policy.

2.12.8 Consideration of factors other than radiograph reading performance

The final supplementary objective was to identify evidence about clinical effects and cost of radiographer reporting plain radiographs, as discussed below, to help justify and inform the design of the primary studies presented in Chapters 4 to 6.

2.12.8.1 The effect of reports on referring clinician's diagnosis

The clarity and certainty conveyed in a radiograph report are highly significant (Audit Commission, 1995; Lafortune et al, 1988). Reports that are complex and equivocal may impede effective communication and confuse the referring clinician (Sierra et al, 1992). Clinicians often place great value on results that do nothing more than reassure them and subsequently the patient (Fryback & Thornbury, 1991). Nevertheless, the effect of radiographer reports on clinicians' diagnosis was not assessed in a single study included in the review.

2.12.8.2 The effect of discordant reports on patient management/outcome

An incorrect report may result in patients receiving further unnecessary radiation or invasive tests, with both financial and health consequences. Only one diagnostic outcome study attempted to assess this by asking A&E consultants to classify radiographers' and casualty officers' false negative interpretations as being clinically important or unimportant [B] [6]. The radiographers' and casualty officers' radiograph reading accuracy was respectively 92.6 per cent (1307/1412) and 92.9 per cent (1331/1432), not significantly different (P=0.66). The percentage of false negative interpretations by radiographers judged to be clinically important was 63.2 per cent (43/68) compared with 54.0 per cent (34/63) for the casualty officers, again not statistically significant (P=0.23). Although there was no significant difference in performance, these findings were only in the context of radiographers red-dotting A&E radiographs. No study provided findings about the effect of radiographer reporting on patient management or outcome.

2.12.8.3 The cost of radiographer reporting

Radiographer reporting could substitute radiologist reporting were it more accurate or equally as accurate at significantly less cost. However, only one study, which evaluated the implementation of radiographer reporting services in four NHS trusts, attempted to conduct a cost analysis [B] [28]. Each NHS trust approached the implementation of the service in its own way. Additional costs identified for providing the service ranged from nil to £15 000 per annum, though it was difficult to ascertain the costs included and how they were analysed. Indeed, the report states that "cost data was not considered to be reliable and more evaluation of costs is required". So despite the CoR (1997) promulgation that "radiographers in a developed reporting role has the capacity to revolutionise the cost-effective management of the patient in clinical radiology" not a single robust economic evaluation had been conducted.

2.13 Conclusions

This review has demonstrated evidence of varying quality that radiographers can report plain radiographs during clinical practice to a high level of accuracy, that is greater than 90 per cent.

There was no evidence to support the claim that radiologists of varying seniority report A&E radiographs significantly more accurately than selectively trained radiographers, although some was provided to support radiographers who have received training to report plain radiographs from sources other than for A&E patients. The results of this review also indicate that training radiographers is an effective strategy for improving their radiograph reporting performance, such that radiographers with training may be able to report images for a variety of modalities.

There was also evidence that radiographers can red dot or triage radiographs to a high level of accuracy, but the quality of the studies suggests that more rigorously designed research is required, particularly to determine whether the combined performance of radiographers and A&E staff can significantly improve accuracy.

Finally, it identified wide variation in the presence of bias and adherence to methodological standards. There was an absence of evidence on the effect of radiographer reports on clinician's diagnosis, choice of patient management or outcome, and the associated costs. These findings were important to inform the design of the primary studies presented in Chapters 4 to 6, and justified these studies assessing the wider clinical effects and costs of radiographer reporting before recommending whether selectively trained radiographers should substitute radiologists in this role. Before these primary studies could be conducted, it was necessary to develop and assess the methods used to ensure that the collection of data when measuring radiographer and radiologist reporting performance was reliable and valid, as discussed in detail in the next chapter.

Chapter 3

Development and assessment of methods for measuring radiographer reporting performance

3.1 Introduction

Chapter 2 presented findings from the systematic review that studies of plain radiographer reporting performance were prone to a variety of biases and that there was a lack of adherence to several methodological standards. In particular, 22 per cent of studies did not use a valid reference standard and nearly all failed to assess whether the criteria for comparing reports for concordance were consistently applied. These findings were surprising, as the complexities in defining a report and the problems with measuring reporting performance, although well known (as discussed in Chapter 1), were largely ignored by studies included in the systematic review.

One of the reasons for identifying threats to the validity of the studies included in the review was to inform the design of the primary studies presented in Chapters 4 and 5. The aim here was to develop and assess the methods used for measuring reporting performance to underpin the validity and reliability of the data collected for the primary studies. A pilot study was conducted to explore how an arbiter, or the person responsible for comparing reports, made decisions and to assess the consistency with which decisions were being made. This resulted in the development of the decision-making criteria to be used in the primary studies of the thesis, which were also assessed for reliability. The acceptability of the reference standard is also discussed for both reliability and validity.

3.2 Objectives

- To develop decision-making criteria for comparing reports for concordance.
- To investigate the acceptability of the reference standard.

3.3 Developing criteria to compare reports for concordance

In order to develop the criteria for comparing reports for concordance in the primary studies, a pilot study was undertaken to explore why an arbiter, or the person responsible for comparing reports, made the decisions that they did. Having developed decision-making criteria for the primary studies, intra- and inter-arbiter reliability was assessed, to ascertain whether the criteria were consistently applied by the same arbiter on separate occasions (intra-arbiter) and by different independent arbiters on the same occasion (inter-arbiter). The Kappa (K) score was used, as this measures agreement rather than percentages. It also takes account of the position of agreement in a contingency table and whether it occurred by chance (Altman, 1991). While no absolute definitions are possible, Table 3.1 provides guidelines to interpret Kappa values.

Table 3.1 Kappa values

Value of K	Strength of agreement
<0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very Good

3.3.1 Pilot study

To develop these criteria, a non-random sample of 25 accident and emergency (A&E) and 25 general practitioner (GP) plain radiographs was selected, stratified 50:50 by the appendicular skeleton and remaining body areas. As described in Chapter 1, this stratification reflects the body areas that the radiographers, or Clinical Specialist Radiographers (CSRs) at Trust A, do and do not report.

The reference standard was an experienced consultant radiologist (as described in Chapter 4) who reported on this quota sample of radiographs using Proforma 3.1. The reference standard had access to previous radiographs when necessary but not to any reports.

Proforma 3.1 Form used by reference standard to record their report

Age		D.O.E		Examination		XR No.	
Sex		D.O.R				Ref No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	
Clinical Details				**Report**			

The author then recorded, on Proforma 3.2, the reference standard report with the original report made during clinical practice. A consultant radiologist at Trust A (i.e. internal arbiter) judged the agreement of the reports and recorded the reason(s) for the decision. The arbiter was blind to who wrote the reports, to clinical details of the patient (e.g. type of patient), and to the reason why the radiograph had been requested. A month later, the arbiter repeated this process, but blind to the original judgements in order to assess whether the arbiter made the same decisions on separate occasions (i.e. intra-arbiter variability). Blinding the reference standard and arbiter was done to avoid the possibility of bias when access to certain information adversely influences results (Sackett et al, 1991; White et al, 1994).

Proforma 3.2 Form used by arbiter to compare reports for the pilot study

		Ref No.	
Original report	**Present report**		
Decision			
Agree			
Disagree			

Table 3.2 shows only two cases when the arbiter's decision changed. The original report for Case 9, an A&E skull radiograph, stated there was 'No sign of trauma', but the reference standard reported 'Soft tissue swelling over the posterior parietal region on the lateral film'. The discrepancy in decision-making occurred because the arbiter changed his mind about the importance of referral to soft tissue swelling as a sign of trauma. The second case was a GP chest radiograph (Case 39). In this instance, the arbiter changed his decision about the relevance of the reference standard statement that it was not possible to exclude the presence of an abnormality since the radiograph was 'underpenetrated'. The test-retest method of measuring the intra-arbiter reliability produced the following Kappa (K) scores when making decisions on A&E radiographs (K=0.90), GP radiographs (K=0.80), and for all cases (K=0.90). Table 3.1 shows that when K is >0.8 this is considered very good agreement.

Table 3.2 Intra-arbiter reliability

Decision	A&E		GP	
	1st Time	2nd Time	1st Time	2nd Time
Agree	18/25	19/25	17/25	18/25
Disagree	7/25	6/25	8/25	7/25

The process was demonstrated to be statistically reliable, but when investigation was made into why the arbiter made his decision some aspects of this were considered inappropriate. First, the arbiter's decision on whether reports agreed was based on how discrepancies might result in different treatment and patient outcome. However, as described in more detail in Chapters 4 and 5, this was for the referring clinician to decide. Second, in the absence of access to clinical details, the arbiter sometimes assumed that the radiograph was of an A&E patient when it was of a GP patient. Without access to this information, the arbiter could make an incorrect assumption about a patient, such as the relevance of a report referring to abnormalities like osteoarthritis. So although access to clinical information can sometimes adversely influence decision-making, the decision was taken to make it available to the arbiter when reports were compared for concordance. Rickett et al (1992) supports the case that clinical details are useful in radiograph interpretation.

3.3.2 *Decision-making criteria used in the primary studies*

A description follows of how reports were compared for concordance in the primary studies based on the findings from the pilot study and further discussion about how this should be done.

Proforma 3.3 was designed to simulate an X-ray request form and to facilitate decision-making. The author recorded the pair of reports to be compared, the reference standard and the observer report. A consultant radiologist at King's Lynn Hospital acted as arbiter in the primary studies. An arbiter external to Trust A would be more objective than an internal arbiter, since his or her own department was not being evaluated. It would also be more difficult for an external arbiter to recognise an individual's reporting style that might indicate whether a report was made by a CSR or radiologist and therefore affect their judgement (Brealey et al, 2002b).

Before deciding whether the reports agreed, the arbiter categorised them as definitely, probably or possibly abnormal, and probably or definitely normal. This was done for two reasons. It reflected the qualifying descriptors used to convey uncertainty as to the presence of an abnormality, which in turn would facilitate the arbiter's decision-making. Second, rather than relying on sensitivity and specificity alone, performance could be presented using receiver operating characteristic (ROC) curves, which exploit an observer's natural tendency for probability scaling when reading radiographs (Manning, 1998). As well as classifying the reports as normal or abnormal, the arbiter could, if necessary, classify them as "equivocal". This provided the option to use this information when analysing the data from the primary studies (Simel et al, 1987). Table 3.3 defines what is normal, abnormal or equivocal.

Proforma 3.3 Form used by the arbiter to compare reports in the primary studies

A&E/GP					
Age		Examination required		Hospital No.	
Sex				Reference No.	
Walking		Ambulance S1/2		Trolley	
Chair		Stretcher		Escort Yes/No	
Date of Examination					

Clinical Details

Report	**Report**

Criteria: Please tick in
The appropriate boxes

Definitely normal	[ ]	Definitely normal	[ ]
Probably normal	[ ]	Probably normal	[ ]
Possibly abnormal	[ ]	Possibly abnormal	[ ]
Probably abnormal	[ ]	Probably abnormal	[ ]
Definitely abnormal	[ ]	Definitely abnormal	[ ]
Normal	[ ]	Normal	[ ]
Abnormal	[ ]	Abnormal	[ ]
Equivocal	[ ]	Equivocal	[ ]

Agree: [ ]
Disagree: [ ]

Table 3.3 Definition of reports

Normal

a	the film is reported as normal e.g. within normal limits, normal for age
b	it is reported that "no abnormality" is present or "No Bony Injury"
c	an "abnormality" is reported but is not relevant in the clinical context the report was made e.g. incidental benign lesions unrelated to the present trauma, healed fractures, degenerative changes which might be expected for the age of the patient

Abnormal

a	a clinically relevant abnormality is present

Equivocal

a	Artefacts/technical defects:	
	i	artefacts - a physiological phenomena (e.g. bowel gas obscuring details, obscured by faeces) or the consequence of other miscellaneous factors (e.g. residue from previous contrast swallow, area of the body covered by parents hand)
	ii	technical defects - as a consequence of poor radiographic technique (e.g. under/over/poorly/well-penetrated films, poor patient positioning, collimating an area of the body)
b	Patient restrictions - poor inspiratory effort, patient severely kyphotic.	
c	Administration - do not have old films for diagnostic comparison, examination needs repeating to be of any diagnostic value	

When comparing reports for concordance it was also important to consider whether an abnormality is present (detection), where is the abnormality (localisation) and what type of abnormality is it (classification) (Robinson, 1996b). The arbiter therefore applied the following criteria when judging whether reports agreed:

- both are categorised as normal, or
- both are categorised as abnormal AND agree on the nature of the abnormality AND agree on the location of the abnormality.

To calculate performance indices such as sensitivity or specificity the observer reports when compared with the reference standard were classified as true positive (TP), true negative (TN), false positive (FP) or false negative (FN).

- TP = the reports agreed on the presence of an abnormality, its location and the type of abnormality.
- TN = the reports agreed on the absence of abnormality.
- FP = the observer report incorrectly identified an abnormal appearance.
- FN = the observer report incorrectly describes the appearances as being normal.
- U = Undecided = the reports agreed on the presence of an abnormality but the location or type of abnormality was incorrectly described.

For the undecided cases the arbiter recorded the reason for disagreement. Using this information the following criteria were applied so that the author could classify the observer report (OR) compared with the reference standard report (RSR) as FN or FP.

- If the RSR recorded the presence of two abnormalities (e.g. A + B) and the OR recorded the presence of only one abnormality (e.g. A) then the OR was classified as FN. This is because OR missed pathology B and was therefore incorrectly more normal than the RSR (i.e. undercalling).

- If the RSR recorded the presence of one abnormality (e.g. A) and the OR recorded that two abnormalities were present (e.g. A + B) then the OR was classified as FP. This is because the OR incorrectly identified an extra abnormality B that the RSR did not (i.e. overcalling).

- If the RSR recorded the presence of abnormality A and the OR abnormality B then the RSR was classified as FN because it missed pathology A. Because the assumption was that the RSR is correct, the emphasis was on the OR not identifying the abnormality (i.e. undercalling).

3.4 Reliability of the arbiters comparing reports

Consistency in the application of the decision-making criteria used in the primary studies to measure reporting performance for both intra- and inter-arbiter reliability is discussed here. The two studies presented in Chapters 4 and 5 used the same sample of radiographs. The first primary study, as described in Chapter 4, compared the original report made in clinical practice with the reference standard, while in the second, described in Chapter 5, the CSR and consultant radiologist reports were compared with the reference standard on separate occasions. The original reports made in clinical practice and used in the first primary study are described as Group B and the CSR and radiologist reports used in the second primary study are described as Group C and D, respectively. To assess arbiter reliability, a 10 per cent (50:50 A&E and GP) convenience sample of Group B, C and D reports was selected from the primary studies.

3.4.1 Assessment of intra-arbiter reliability

To measure intra-arbiter reliability, the external arbiter repeated the process of applying the decision-making criteria described in section 3.3.2 to the sample of reference standard versus Group B, C or D reports six months later. Table 3.4 shows that the external arbiter achieved Kappa scores of greater than 0.6, which is good agreement, when comparing the same sample of reports on two separate occasions for all groups of observers and the different patient types. Some Kappa scores even exceeded 0.8 - very good agreement. These Kappa scores indicate that the external arbiter was consistent in applying the decision-making criteria.

Table 3.4 Intra-arbiter reliability

Group	Type	n	(%)	Kappa
B	A&E	35/40	(88)	0.66
	GP	37/40	(93)	0.89
C	A&E	34/40	(85)	0.67
	GP	32/40	(80)	0.70
D	A&E	37/40	(93)	0.82
	GP	35/40	(88)	0.82

3.4.2 *Assessment of inter-arbiter reliability*

To measure inter-arbiter reliability, one of two consultant radiologists at Trust A (i.e. internal arbiters) made an independent comparison of the 10 per cent sample of reports from the primary studies using these criteria. The decisions made by the internal arbiters were then compared with the *first* and *second* decision made by the external arbiter.

Inter-arbiter variation is usually greater than intra-arbiter variation (Brealey & Scally, 2001). Table 3.5 shows the same applies to this study, with the Kappa scores for inter-arbiter agreement generally lower than the scores in Table 3.4. Although there are two outliers (K = 0.40 and 0.96), the Kappa scores are around 0.6 or more, indicating good agreement.

Table 3.5 Inter-arbiter reliability

Group	Type	First decision			Second decision		
		n	%	Kappa	n	%	Kappa
B	A&E	32/40	(80)	0.57	30/40	(75)	0.40
	GP	28/40	(70)	0.55	29/40	(73)	0.60
C	A&E	35/40	(88)	0.75	33/40	(83)	0.63
	GP	31/40	(78)	0.66	39/40	(98)	0.96
D	A&E	35/40	(88)	0.70	32/40	(80)	0.55
	GP	29/40	(73)	0.60	30/40	(75)	0.63

3.4.3 *Conclusion about arbiter reliability*

This section provides evidence of good agreement in terms of both intra- and inter-arbiter reliability when comparing reports for concordance using the decision-making criteria developed for the primary studies presented in Chapters 4 and 5. Nevertheless, some disagreement arose between arbiters as to whether reports were concordant. Chapter 2's finding, that the diagnostic threshold explained some of the variation in the results of studies using *different* samples of radiographs, is not surprising in light of this variation in decision-making when using the *same* sample of radiographs.

3.5 Acceptability of the reference standard

The findings discussed in section 3.4 suggest that the decision-making criteria were applied consistently in both primary studies, but this does not ensure that the results of the study will be reliable and valid, as the choice of reference standard is another salient, contributory factor. For the two primary studies (as described in Chapters 4 and 5) an experienced consultant radiologist provided the expert, or reference standard report. All the radiologists at Trust A agreed that the consultant was an acceptable departmental standard, but it was still essential to provide evidence to support their opinions. The following describes the assessment of the reference standard for reliability and validity.

3.5.1 Assessment of the reliability of the reference standard

The consultant radiologist acting as reference standard was asked to report on the same sample of A&E and GP radiographs used in the pilot study eighteen months earlier. A previous study assessing observer agreement between five consultant radiologists used a five month delay between the first and second viewings (Tudor et al, 1997).

The reference standard was again provided with Proforma 3.1 to record his report and had the same information available to him as in the pilot study (section 3.3.1). The author then recorded the first and second reference standard reports on Proforma 3.3, so that the external arbiter could judge concordance. Table 3.6 shows *very good* agreement between the reference standard reports for A&E radiographs and *good* agreement for GP radiographs and all radiographs in total, a considerably better result than was found by Tudor et al (1997). They showed that the Kappa values for agreement between five consultant radiologists when reporting plain radiographs improved with access to clinical information from 0.31 to 0.58.

Table 3.6 Reliability of the reference standard

Type	n	(%)	Kappa
A&E	23/25	(92)	0.85
GP	20/25	(80)	0.65
Total	43/50	(86)	0.75

Reference standard reports were discordant for seven radiographs (i.e. two A&E and five GP), explicable, for example, by disagreement on the number of lumbar spine vertebrae for which degenerative disease was present. Reports agreed, then, on the presence and type of abnormality but not its location. The arbiter confirmed that none of the discrepancies between these reports would have resulted in a clinically important difference in patient management or outcome, so that even when variation occurred in the reference standard reports, it was not clinically important.

As a further measure of the acceptability of the intra-observer agreement for the reference standard, the intra-observer agreement between the CSRs and consultant radiologists was calculated. This was possible because the same sample of radiographs were included in Chapters 4 and 5, enabling a comparison, for example, between the CSR report of a radiograph made during clinical practice (see Chapter 4), and a report of the same radiograph by the same CSR (see Chapter 5). The CSRs intra-observer agreement, for A&E radiographs of the appendicular skeleton only, was good (K=0.77, n=42), but the radiologist intra-observer agreement in total was poor (K=0.40, n=53). Higher Kappa values were achieved by the reference standard than either the CSRs or consultant radiologists, providing further evidence of the report's reliability, though the different sample of radiographs could explain this variation in intra-observer agreement.

3.5.2 Assessment of the validity of the reference standard

As discussed in greater detail in Chapter 4, patient outcome was assessed by investigating whether erroneous A&E radiograph reports made during clinical practice resulted in patient re-attendance to the A&E or X-ray department.

The results of the study presented in Chapter 4 show no patients re-attending the A&E department, although there were eight cases when a patient returned to the X-ray department for a further radiograph of the same anatomical area and the follow-up radiograph report suggested inaccuracy in the index report. For the purpose of assessing patient outcome in Chapter 4, three consultant radiologists acted as a 'gold standard' and judged whether an error had been made in the original report of these eight cases. To validate the reference standard, these same consultant radiologists also judged the accuracy of his report.

The gold standard agreed that the reference standard was correct in seven of the eight cases. For Case 765, a radiograph of the elbow, the reference standard suggested there was a 'Small bony flake adjacent to the lateral epicondyle', whereas the gold standard judged this to be a 'Fleck of calcium'. This slight difference in the interpretation of the radiograph would not be clinically important (De Lacey et al, 1980). Furthermore, both reports agreed that there was 'no convincing radial head fracture'.

3.5.3 Conclusion about the acceptability of the reference standard

The empirical evidence presented here supports the radiologists' opinion at Trust A that the chosen consultant radiologist would provide an acceptable reference standard report. Even when the reference standard reports on the same radiographs were shown to disagree with each other, or the reference standard differed from the 'gold standard', these differences were judged not to be clinically important.

3.6 Conclusion

Chapter 3 presents evidence that the methods developed for measuring reporting performance should underpin the validity and reliability of the data collected for the primary studies in this thesis. Chapter 4 describes the first of these primary studies, which assesses the introduction of selectively trained radiographers to the A&E reporting service at Trust A, in terms of radiograph reporting accuracy, patient management and outcome.

Chapter 4

Introducing selectively trained radiographers to an A&E reporting service: A retrospective controlled before and after study

4.1 Introduction

Chapter 1 discussed skill-mix initiatives as an opportunity to lighten radiologists' workload, whilst increasing the scope, challenge and interest of other staff. Findings on radiographer performance when reporting plain radiographs were presented in Chapter 2. Chapter 3 discussed the methods for measuring reporting performance for the primary studies in Chapters 4 and 5.

The aim of the study presented in Chapter 4 is to assess the introduction of selectively trained radiographers, or Clinical Specialist Radiographers (CSRs) as they are now called, reporting accident and emergency (A&E) radiographs of the appendicular skeleton at Trust A. Its rationale was that at the time of its conception the CSRs at Trust A had been reporting A&E plain radiographs of the appendicular skeleton for two years and the staff of the X-ray Department wanted to determine whether they had been successfully introduced without detriment to the reporting service.

The result of background research for the systematic review in Chapter 2 indicated that according to methodological papers on evaluating diagnostic tests, studies of radiographer plain radiograph reporting performance were of varying quality. Issues concerning the effect of reports on clinician's diagnosis, patient management and outcome and the associated costs were also consistently ignored, so it was deemed timely to design and conduct a study that adhered to the methodological standards that should underpin the assessment of radiographer reporting performance and to eliminate, when feasible, all sources of bias.

Absence of evidence on the clinical effects and costs of radiographer reporting meant that it was also important to design a study that addressed these issues, a requirement highlighted by the College of Radiographers' (1997) promulgation that radiographer reporting could revolutionise the cost-effective management of the patient in clinical radiology and other imaging dependent services.

Therefore, the *primary* objective of this study was to assess the effect on A&E radiograph reporting accuracy, of introducing selectively trained radiographers, or CSRs, at Trust A. The *secondary* objectives were to assess the subsequent effect of introducing CSR reporting on patient management and outcome, and the associated costs. To achieve this, only incorrect reports were used to evaluate whether they had a detrimental effect on patient management and outcome.

In order that the findings should be a valid representation of the consequences of introducing these radiographers to clinical practice, a pragmatic study design was employed. As an example, a retrospective sample of reports was collected, so that when the radiographs were reported, the different professionals had access to all the information they would normally have (e.g. previous radiographs and reports). The referral would have been presented to them using a standard X-ray request form and the radiographs would have been reported under normal conditions, including the CSRs discussing radiographs with radiologists as necessary. The exploratory analysis of the cost of introducing CSR reporting undertaken alongside this study is presented in Chapter 6.

4.2 Objectives

4.2.1 Primary objectives

1. To assess the effect of introducing selectively trained radiographers on A&E radiograph reporting accuracy.

4.2.2 Secondary objectives

2. To assess the effect of introducing selectively trained radiographers reporting A&E radiographs on patient management.

3. To assess the effect of introducing selectively trained radiographers reporting A&E radiographs on patient outcome.

4.3 Setting

This section describes the A&E reporting systems used at Trust A, which consists of two clinical sites (A and B), the point of referral for radiographs for the CSRs and radiologists to report.

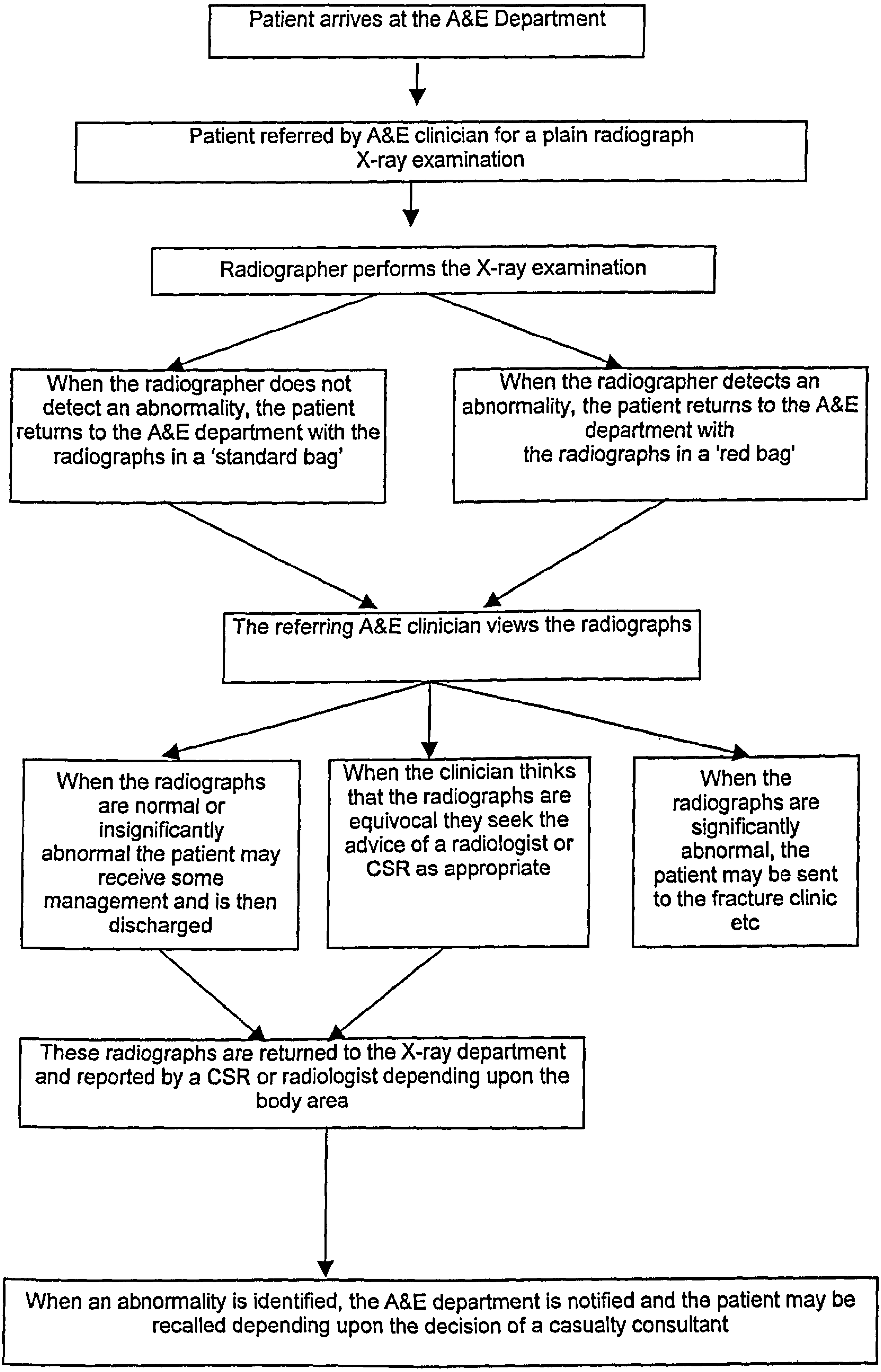
4.3.1 A&E reporting system at site A

In the first instance, casualty officers refer A&E patients to the X-ray department for a radiographic examination. When radiographers performing the X-ray examination notice an abnormality on the radiograph, the patient is asked to return to A&E with the radiographs in a 'red bag', indicating to the referring casualty officer that an abnormality was identified. If the casualty officer judges that the abnormality is clinically significant, the patient is sent to the fracture clinic or treated appropriately. This may include returning for a follow up.

In the event that the radiographer who performed the X-ray examination judges the radiograph to be normal, the patient is asked to return to A&E with the radiographs in a 'standard bag'. If the casualty officer agrees that the radiograph is normal or detects a clinically unimportant abnormality, the patient may receive some management and is then discharged. The radiographs are returned to the X-ray department to be reported, depending on the body area, by a CSR or radiologist. When the radiograph judged to be normal by the radiographer who took the X-ray is considered abnormal by the casualty officer then the patient is treated appropriately. A casualty officer who is equivocal about the presence of an abnormality will visit the X-ray department to ask for the opinion of a CSR or radiologist as appropriate.

The A&E radiographs judged normal by both the radiographer who performed the X-ray examination and the referring casualty officer are reported the next day by a CSR or radiologist. On detection of an abnormality, the A&E department is notified and the patient may be re-called depending upon the decision of a casualty consultant. Currently, the two CSR report A&E radiographs on alternate weeks. Figure 4.1 overleaf illustrates the reporting system at site A.

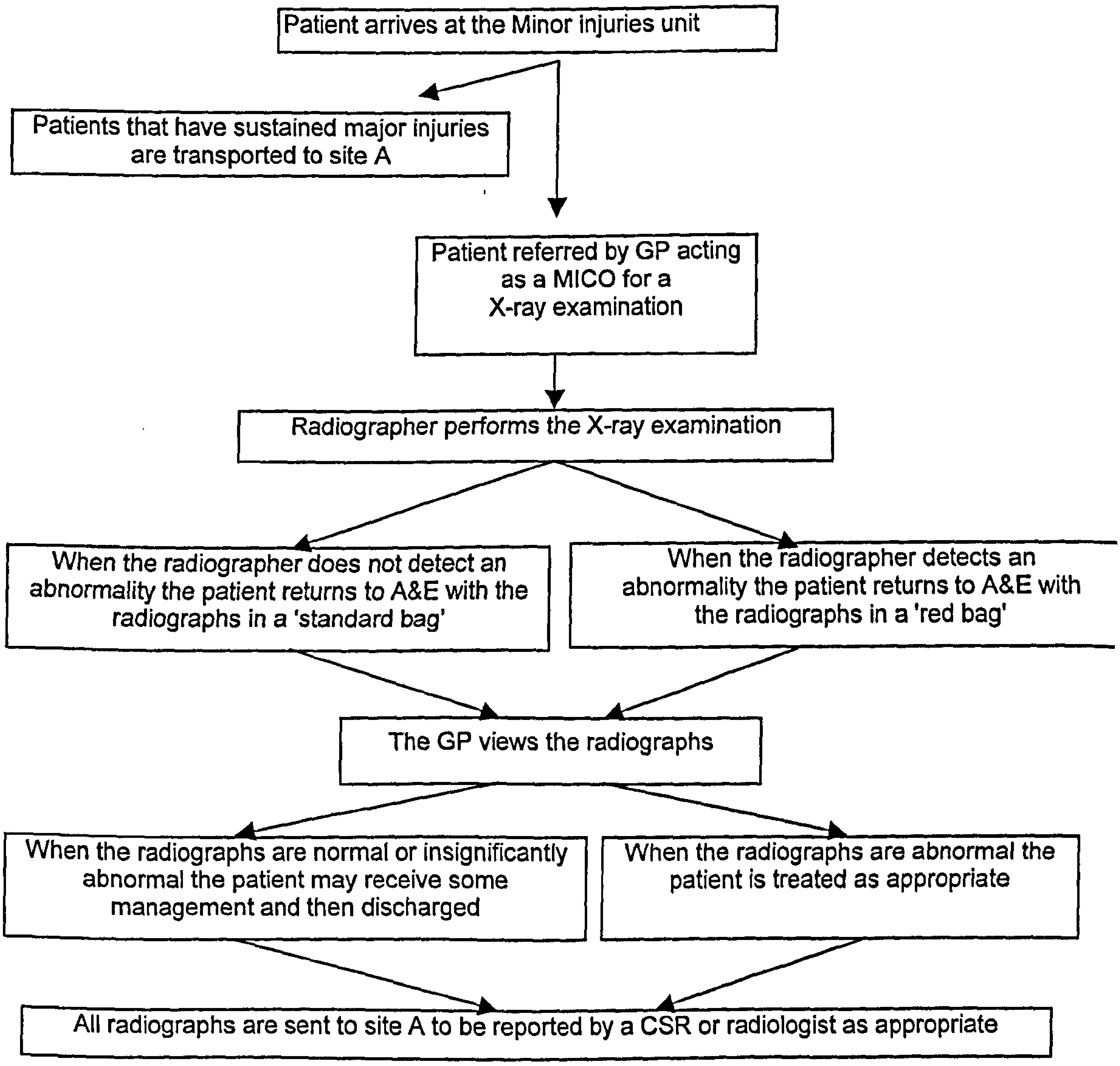
Figure 4.1 Site A reporting system



4.3.2 A&E reporting system at site B

Very few patients attend the minor injuries unit at site B, and those who experience major injuries are referred to site A. General Practitioners (GPs) acting as minor injury casualty officers (MICO) decide whether a patient should be given an X-ray examination. When the radiographer who performs the X-ray examination locates an abnormality the patient is returned to the GP with the radiographs in a 'red bag'. Those with examinations judged normal are returned to the GP with the radiographs in a 'standard bag'. The GP then assesses whether the examination is abnormal and how the patient should be managed. All examinations are sent to site A to be reported by a CSR or radiologist as appropriate. Figure 4.2 presents the A&E reporting service at site B.

Figure 4.2 Site B reporting system



4.4 Methods

The purpose of this section was to discuss the methods used to assess the effect of introducing selectively trained radiographers on A&E radiograph reporting accuracy, and the subsequent effects on patient management and outcome.

4.4.1 Assessing radiograph reporting performance

The study's primary objective was to assess the effect of introducing selectively trained radiographers on A&E radiograph reporting accuracy, providing information on whether abnormalities were being missed (i.e. undercalling) or the presence of an abnormality was reported when the radiograph was normal (i.e. overcalling). When professionals 'undercall', it results in a high number of false negatives and low sensitivity, such that patients do not receive the treatment they require, which can affect patient outcome. 'Overcalling' results in a high number of false positives and low specificity, and patients may receive treatment inappropriately and experience unnecessary anxiety. Both 'undercalling' and 'overcalling' also have resource implications.

In February 1995, following a period of training to report all body areas and different patient types, two selectively trained radiographers, or CSRs, began reporting plain radiographs referred from the A&E department at site A and B. They have been reporting the appendicular skeleton i.e. upper limb (shoulder girdle to fingers), lower limb (hip to toes) and foreign body examinations, but *not* the axial skeleton (pelvis, spine, head), chest, thorax and abdomen. The timing of this intervention and the type of patient and body areas that the CSRs and radiologists report was important for informing the design of the study as discussed below.

The first important factor to consider was the time period for data collection. The CSR began reporting A&E radiographs of the appendicular skeleton in February 1995. Up until that date, radiologists made these reports. Over the next 18 months, there were different phases to their introduction, affecting whether the CSRs used codes only or a combination of codes and free text to report radiographs, and whether they were allowed to report only normal or both normal and abnormal radiographs. So not all A&E appendicular skeleton radiographs were reported by the CSR - some were still reported by radiologists. Data was, therefore, collected retrospectively from February 1993 to January 1997, two years before and after the introduction of CSR reporting.

Data collection over a four-year period was important for several reasons. This period covered all the different phases of the introduction of CSR reporting. Furthermore, it was important to minimise *statistical regression.* A sample of reports might have been collected, for example, only one month before and after the introduction of CSR reporting. If it was found that performance increased subsequent to their introduction, this could be a genuine finding, but it could also be explained by random factors like a change in the case-mix of patients, or staff being ill or on holiday. The performance after the introduction of the intervention may simply have been a result of regression towards the grand mean. Data collection over a long period would represent a more stable and more valid reflection of performance. It was also essential to consider the effect of *maturation.* For example, as the CSRs gained experience and confidence in reporting, these 'learning effects' may have resulted in improved performance over time, and data collection over a short period may not have accurately reflected this.

Selecting reports over a long period would enhance the internal validity of the study, but between February 1993 and January 1997 around 97,000 A&E plain radiograph examinations were performed. Assessing the accuracy of a report requires comparison with a reference standard and it was clearly not feasible for a reference standard to interpret this number of examinations. A smaller but representative sample of radiographs needed to be selected for the reference standard to report, for comparison with the original report made during clinical practice. This was achieved by selecting a random stratified sample (as discussed later).

Another key methodological factor to consider was the choice of control group. One possible control was a random stratified sample of A&E reports, selected before the introduction of CSR reporting, with a further random stratified sample selected after their introduction as the experimental group. This design is presented below, where O is the observation, X the intervention, and the subscript refers to the sequential order of recording observations, described as a 'one group before and after design' by Cook & Campbell in 1979.

A&E appendicular skeleton	O_1	X	O_2

If a comparable sample of A&E appendicular skeleton reports was selected before and after the introduction of CSR reporting and performance was found to be higher after their introduction, it might be reasonable to attribute the change to the intervention, but the change might alternatively be due to *history,* in that other events could have affected performance. Notably, even after the CSRs began reporting, some A&E appendicular skeleton radiographs were still reported by radiologists. A change in radiologists at Trust A before and after the introduction of CSR reporting could possibly explain a change in performance, making it necessary to rule out any plausible historical factors that could affect performance when drawing conclusions.

An alternative study design, to help control for historical events, was the selection of a sample of A&E appendicular skeleton (upper limb, lower limb) reports as the experimental group and the A&E axial skeleton (pelvis, spine, head) reports as the non-equivalent, non-intervention control group. The A&E axial skeleton could be described as non-equivalent because of its differing range of body areas from the appendicular skeleton, but it is still similar to the experimental group, as these radiographs are for the same type of patient and often for similar pathology (e.g. fractures). Since the CSR do not report A&E axial skeleton radiographs, this is a non-intervention control group.

A&E appendicular skeleton	O_1	X	O_2
A&E axial skeleton	O_1		O_2

One concern with this design was the possible effect of introducing CSR reporting on the radiologist's performance. Suspicion of fracture is one of the main reasons for referral for A&E radiograph of the skeleton. At Trust A, the rota comprised around seven radiologists for reporting A&E radiographs. The introduction of CSRs could have a considerable effect on the number of A&E radiographs of the skeleton they would individually report and consequently on their ability to recognise fractures. So although the A&E axial skeleton is a non-intervention control group, the introduction of CSR reporting could have an adverse effect on the accuracy with which these radiographs are reported by radiologists.

To address these possible threats to the validity of the study, a random stratified sample of A&E plain radiograph reports before the introduction of CSR reporting was used as one control group and a random stratified sample of A&E plain radiograph reports *after* its introduction was the experimental group. In addition, a random stratified sample of GP reports was included before and after the introduction of CSR reporting as a non-equivalent, non-intervention control group. (These reports are non-equivalent because the GP and A&E referrals are for a different type of patient with different pathology, generally degenerative problems). Assessing reporting performance over time for GP patients is still valid, as in the case of A&E patients, they were being referred for plain radiograph X-ray examinations. It is a non-intervention control group as the CSRs do not report GP radiographs. A reduction in the number of A&E radiographs that the radiologists report should not affect their ability to report GP plain radiographs.

A&E	O_1	X	O_2
GP	O_1		O_2

Hence, the final choice of design was a retrospective controlled before and after study design, including the selection of a random stratified sample of A&E and GP reports for all body areas. Unlike the traditional use of such a control group, there was no intention to test for a difference in accuracy between A&E and GP reports before CSR reporting and make the same comparison afterwards. Instead, it represented an additional control group to help further underpin the study, making it possible to correctly infer that any change in A&E radiograph reporting performance could be attributed to the CSRs. Sub-group analysis of A&E radiograph reporting accuracy for the appendicular skeleton separate from the remaining body areas was important to ensure the latter was not masking differences in performance.

4.4.2 What sample size was necessary?

Consultant radiologists were reporting A&E radiographs before the introduction of the CSRs. Therefore, evidence was required to calculate the sample size needed to detect both statistically and clinically significant differences in accuracy before and after introduction of the CSRs.

The reporting performance of two selectively trained radiographers at St James' University Hospital who attended the same training program as those at Trust A has already been assessed (Robinson, 1996a). Between them, the Leeds CSR reported 561 A&E musculoskeletal, chest and abdomen plain radiographs in parallel with rota radiologists of varying seniority. The study found the radiographers and radiologists respectively reported at 95.7 per cent and 97.1 per cent accuracy. Whereas Robinson (1996a) compared the radiographers to radiologists of varying seniority this study was comparing the CSR to consultant radiologists, so the difference between the two groups of professionals should be greater in this study. It seemed reasonable to assume that if the CSR reported 5 per cent less accurately than the radiologists this difference would be clinically important. Consequently, the null hypothesis assumed no significant difference in the accuracy with which A&E radiographs are reported before and after the introduction of the CSR. The alternative hypothesis assumed a 5 per cent difference in accuracy before and after the introduction of CSR reporting and detection of this difference would be clinically significant.

Table 4.1 shows, using nQuery 4.0, that 326 A&E plain radiographs would be required in each sample (or 652 in total) before and after the introduction of CSR reporting to have 80 per cent power to reject the null hypothesis when the alternative hypothesis is true at the 5 per cent significance level. However, there was only a short period of time during which this assessment of accuracy could be performed and the subsequent influence of incorrect reports on patient management and outcome. Table 4.1 also shows the power of the study for the number of radiographs that could realistically be included in the study, two hundred A&E radiographs before and after the introduction of CSRs (or 400 in total). This sample size would have 59 per cent power to reject the null hypothesis and accept the alternative hypothesis that there is a statistical and clinical important difference in A&E radiograph reporting accuracy before and after the introduction of CSR reporting. Finally, Table 4.1 presents the power of the study for a 10 per cent significance level.

The quite low power to detect the expected difference in accuracy means the results of the study should be interpreted with caution. Nevertheless, it was still useful to conduct the study, in that Trust A could be provided feedback as to the effect of introducing CSR on the accuracy with which A&E radiographs were being reported. The study was also needed to assess the subsequent effect of incorrect reports on patient management and outcome, which at the time of study design there was no evidence.

Table 4.1 Sample size calculations

Test significance level, α	0.050	0.050	0.010	0.010
1 or 2 sided test	2	2	2	2
A&E proportion before	0.970	0.970	0.970	0.970
A&E proportion after	0.920	0.920	0.920	0.920
Odds ratio	0.356	0.356	0.356	0.356
Power (per cent)	80	59	80	70
N per group	326	200	257	200

4.4.3 Eligibility criteria

A&E and GP plain reports were eligible for inclusion unless the original report was not available from the patient's radiograph bag or the computer system. Reports made during clinical practice by the same person who was to provide the 'reference standard' report were excluded to prevent incorporation bias (Brealey & Scally, 2001).

4.4.4 Study sample of reports

A retrospective random stratified sample of A&E reports was selected from February 1993 to January 1997 (two years before and after the CSRs began reporting). This was an attempt to ensure the sample was comparable before and after their introduction and that it reflected the population of reports from clinical practice.

It was important to stratify the sample by two potential confounding factors: time and body area. For the former, each year was divided into quarters as, for example, different types of fracture occur in different patient groups at different times of the year (greenstick fractures in children during summer and fracture neck of femur in an older population during winter). The sample was stratified by body area so that the same proportion of A&E appendicular skeleton reports, which are the body areas that the CSR report, was included in the sample before and after they began reporting. There was also evidence that the appendicular skeleton is less difficult to report than other body areas (Renwick et al, 1991; Robinson et al, 1999). So the sample of reports selected each quarter was stratified by these body areas.

Random selection of the sample of A&E reports took place in two stages. First, each day of each month was numbered for each quarter (i.e. from one to 89 or 92 depending on the number of days per quarter). A uniform sampling fraction of one in thirty was used. Using a random numbers table, three days per quarter were selected producing a proportionate stratified random sample of days (Moser & Kalton, 1979). Second, the Systems and Networks Service Department at Trust A provided a list of all the A&E examinations performed for the selected days at sites A and B, which acted as the sampling frame. For each quarter the eligibility criteria were applied and each eligible A&E examination was stratified by body area as two separate lists and numbered sequentially. Using a random numbers table, 25 A&E reports were selected per quarter: 50 per cent for each of the two body areas. Because a different number of A&E plain radiographs were performed each quarter but the same number of A&E reports was selected, a disproportionate stratified random sample was produced. So although each A&E report had an equal chance of being selected per quarter, each A&E report within the overall time period did not have an equal chance of being selected. A multi-stage sampling design was necessary to build in the appropriate representation of reports in the sample by not leaving it to chance - a method repeated to select a random stratified sample of GP reports.

4.4.5 Method of reporting by the reference standard

The assessment of the acceptability of the reference standard used in this study was discussed in Chapter 3: a single consultant radiologist at Trust A who had eleven years experience in Radiology and a special interest in skeletal radiology. The radiologist reported all eight hundred radiographs in normal viewing conditions using the same codes and free text (i.e. a written component) as during clinical practice. He had access to previous radiographs, but the original and previous reports were removed from the patient's bag to prevent 'reference standard review bias', as previous reports could prejudice the reference standard report (Brealey & Scally, 2001). To assist the radiologist reporting of radiographs, Proforma A4.1 was designed like an X-ray request form. This proforma is presented in Annex 4, along with all the other proformas referred to in this chapter.

The sample of radiographs was stored in the radiologist's room. Each radiograph bag had a label attached to it with one or more reference numbers written on it ranging from one to eight hundred and kept in numerical order. On the top right-hand corner of each proforma was a reference number corresponding to the number on the radiograph bag. The reference standard recorded the report on the proformas and the author then used a table in Word to record the reference standard report, original report made during clinical practice, and initials of the person who made the report.

4.4.6 Method of comparing reports

Chapter 3 discussed development of the marking criteria so that a health care professional, or arbiter, could judge whether reports were concordant. The arbiter who applied these criteria was a single consultant radiologist at a different hospital with ten years experience in Radiology and also with a special interest in skeletal radiology. The author recorded both reports on Proforma A4.2 so that the arbiter, blind to who produced each report, could compare them for concordance. The arbiter categorised reports as definitely, probably or possibly abnormal, and probably or definitely normal. Normal was defined as within normal limits, no bony injury, or a clinically unimportant abnormality such as healed fractures. Abnormal was defined as all clinically relevant abnormalities. Reports were only judged concordant if they agreed on the presence, location and type of abnormality (Robinson, 1996b).

4.4.7 Assessing effect on patient management

This section describes the method used for addressing the secondary objective of this study, concerning the effect of introducing selectively trained radiographers on patient management. It includes assessment by the A&E Specialist Registrar at St James' Hospital, Leeds and a GP at Malton general practice of the effect on patient management of reports found to be discordant to the reference standard. Knowledge that only discordant reports were used might affect the clinician's decision about an expected difference in patient management (Sackett et al, 1991). So for every two discordant reports a concordant report was included (i.e. 2:1 ratio). The clinician was blind to which reports were discordant. Concordant reports were used to prevent expectation bias, so only a convenient sample was included.

4.4.7.1 Effect on A&E patient management

An A&E staff nurse at Trust A collected the case notes of relevant patients and their clinical details and management were recorded on Proforma A4.3 before the original report was available. The staff nurse also used the re-call note book, case notes, and the casualty cards to identify patients who had been recalled within one month of the original examination. Casualty cards with the same name as these patients were also searched to determine whether the re-called patient had been given a new card. If the patient was recalled subsequent to the original report, the management they received was recorded on Proforma A4.3. The staff nurse did not know which patients' reports were discrepant, so it would not affect the rigor of the search.

The author then recorded the information on Proforma A4.3, except whether or not the patient was recalled, on to Proforma A4.4. Using this information, the reference standard report, and the radiographs when necessary, the A&E Specialist Registrar recorded whether the patient would have been recalled using the options listed in Table 4.2. The clinician considered the potential anxiety experienced by the patient if they were to be recalled and any relevant medico legal issues. When completing Proforma A4.4 the Registrar was blind to the content of original report, whether or not the patient had been recalled, and that it was the reference standard report. Such information could inappropriately affect their judgement.

Table 4.2 Management options

	Management	Example
1.	Admit patient	Compound fracture
2.	Fracture clinic	Closed Colles fracture
4.	Dressing clinic	Bad ankle sprain
4.	GP review	PIPJ volar plate fracture
5.	Refer to a specialist centre	Neuro
6.	TRIN (To Return If Necessary)	Lateral malleolus flake
7.	Discharge	No follow up/normal
8.	Specialist clinic (medical)	Ear Nose Throat (ENT)
9.	Specialist clinic (non-medical)	Physiotherapy
10.	Injury clinic	Review by A&E consultant
11.	Advice (verbal/printed sheet)	No bony injury/soft tissue swelling

Once the A&E Senior Registrar had completed Proforma A4.4, the author recorded the clinical details of the patient and their management, before a report was available, onto Proforma A4.5. Decisions as to whether the patient should be recalled, based on the reference standard and original report, were also included, as were both reports. Using this information and blind to who made which report, the Registrar chose one of the three following responses and recorded the reason for the choice:

- *No difference in management (treatment or advice).* The patient would not have been recalled due to either report or the patient would have been recalled due to both reports and received the same management.
- *A clinically unimportant difference in management.* The patient would have been recalled as a consequence of both reports but any difference in patient management would not affect patient outcome.
- *A clinically important difference in management.* The patient was recalled based on one report but not the other or the patient would have been recalled due to both reports and the difference in observed and expected management would affect patient outcome.

4.4.7.2 Effect on GP patient management

Unlike assessing the effect of reports on A&E patient management, which used data from clinical practice, it was only possible to hypothesise how reports may have affected GP patient management. To do this, a GP completed Proforma A4.6 by recording the expected management of the patient using the clinical details from the relevant X-ray request form and the original report and later repeated the process using the reference standard report. On both occasions, the GP was blind to who made which report. The GP then completed Proforma A4.7, which included the clinical details, the original and reference standard report and the respective expected management. Using this information and blind to who made which report, the GP compared the expected management and then chose one of the three following responses, recording the reason for their choice:

- *No difference in the expected patient management (treatment or advice).*
- *A clinically unimportant difference to the expected patient management.*
- *A clinically important difference to the expected patient management.*

4.4.8 Assessing effect on patient outcome

The following method is that used for addressing the secondary objective of this study about the effect of introducing selectively trained radiographers on patient outcome. Patient re-attendance to the A&E or X-ray department was used as a proxy for patient outcome, because re-attendance indicated that significant pathology might have been missed in the original report.

With the same discordant and concordant reports described in the previous section, the A&E staff nurse used the casualty X-ray card and patient case notes to record on Proforma A4.3 whether a patient re-attended the A&E department within three months of the initial investigation. Clinicians involved in the study judged this an acceptable time to expect a patient to re-attend A&E for a related incident. The Radiology department's computer management system was also searched to establish whether a patient re-attended for further radiographic examinations or other procedures one year from the date of the initial examination. This information was used to identify the reason why a patient re-attended, as described below.

4.4.8.1 Patient did not re-attend A&E or X-ray department

If, after checking the casualty card and case notes and searching the Radiology computer system, it was found that a patient had not re-attended either the A&E or X-ray Department at Trust A, it would suggest that even if the original report was reported inaccurately, there was negligible effect on patient outcome as the patient did not seek further health care from these departments. These patients were not followed up any further.

4.4.8.2 Re-attended A&E department and original radiographs reviewed

This indicated that a patient re-attended A&E for a reason related to the original X-ray examination but only the original radiographs were reviewed. If this occurred it would be important to ascertain whether the original radiographs were found to be erroneously reported, which may have contributed to why the patient re-attended A&E.

4.4.8.3 Re-attended for unrelated X-ray examination

This option was recorded if a patient re-attended for a further X-ray examination unrelated to the same symptomatic body area as for the original injury. Patients in this category would not be followed-up further, as it was assumed that their re-attendance at the X-ray department was for an unrelated reason.

4.4.8.4 Re-attended for related X-ray examination

This was recorded when patients re-attended for a repeat X-ray examination of the same anatomical area or for a different procedure (e.g. CT). It was important to establish whether an error had been made in the original report which resulted in patient re-attendance for another X-ray examination.

A consultant radiologist at Trust A compared the original with subsequent reports to establish whether the reason for re-attendance was an erroneous original report. Should discrepancies be noted in this comparison, or if the subsequent report could not be compared with the initial one, or if the outcome of a different examination raised suspicion as to the accuracy of the original report, all examinations were reviewed in consensus by a group of radiologists, to establish whether the original report was indeed erroneous. The group comprised three consultant radiologists. It did not include the reference standard radiologist or a radiologist responsible for the report of radiographs under review, who were excluded to eliminate their potential for influencing the consensus or 'gold standard' report. If the three consultant radiologists confirmed the initial report was erroneous, the reason was recorded (e.g. occult fracture diagnosed on the second visit, fracture missed on the initial visit, or false positive report on the initial visit).

4.4.9 Data analyses

Radiograph reporting performance was presented based on sensitivity, specificity and accuracy. Stats Direct was used to calculate the approximate (Wilson) 95 per cent confidence intervals (CIs) around a single proportion and the 95 per cent CIs when testing for a difference in proportions between two independent groups (i.e. before and after samples) and the appropriate hypothesis tests (Chi-square test or Fisher's Exact test if the total number of observations were less than twenty or any of the expected frequencies were less than five (Altman, 1991)).

Stats Direct was also used in analysis of patient management data to calculate the 95 per cent CI when testing for a difference in proportions before and after the introduction of CSR reporting. P-values for these tests are not presented. Small sample sizes resulted in low power for detecting a statistically significant difference, so relying on P-values could be misleading. Emphasis was placed on interpreting the 95 per cent CI to indicate precision of the estimated difference in proportions.

Results of the effect of erroneous reports on patient outcome were presented using descriptive statistics.

4.5 Results

Results of the effect of introducing selectively trained radiographers on A&E radiograph reporting accuracy are presented here, followed by the results of the subsequent effect on patient management and outcome.

4.5.1 Effect on radiograph reporting performance

Tables 4.3 to 4.6 present findings about the effect of introducing CSRs on radiograph reporting performance with the A&E appendicular skeleton radiographs described as A&E1 and the remaining body areas as A&E2. Respective body areas for GP patients are described as GP1 and GP2. The total number of A&E1 radiographs that CSR reported was 66/101 (65 per cent).

The main finding from Table 4.3 is that whereas specificity is high, in the nineties, sensitivity is only around 50 per cent, indicating that normal radiographs were correctly being identified but nearly 50 per cent of all abnormal radiographs were being missed (i.e. undercalling). There is also more variation in sensitivity before and after the introduction of CSR reporting than there is for specificity. Table 4.3 shows a fall in specificity by 3 per cent or 4 per cent depending on the body area, but an 8 per cent increase in sensitivity for the reporting of A&E1 radiographs and a fall of 13 per cent for A&E2 radiographs. Overall, however, there was only a one per cent fall in accuracy.

Table 4.4 shows no significant difference in A&E radiograph reporting performance for the two subgroups and in total. It also shows, for example, that for A&E1 radiographs there is 95 per cent certainty that reporting accuracy was between 10.1 per cent lower and 8.3 per cent higher after CSR reporting was introduced (P=0.84). Although the P-value indicates there is no evidence to suggest a significant difference, the CIs are quite wide, reflecting the low power.

Table 4.3 *A&E plain radiograph reporting performance*

BEFORE	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
A&E1	12	77	10	1	100	55 (35, 73)	99 (93, 100)	89 (81, 94)
A&E2	11	73	13	3	100	46 (28, 65)	96 (89, 99)	84 (76, 90)
Total	23	150	23	4	200	50 (36, 64)	97 (94, 99)	87 (81, 91)
AFTER	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
A&E1	15	74	9	3	101	63 (43, 79)	96 (90, 99)	88 (80, 93)
A&E2	5	77	10	7	99	33 (15, 58)	92 (84, 96)	83 (74, 89)
Total	20	151	19	10	200	51 (36, 66)	94 (89, 97)	86 (80, 90)

Table 4.4 *Test for difference in A&E plain radiograph reporting performance*

Statistic	Body area	% Difference (95% CI)	P-value
Sensitivity	*A&E1*	8 (-20.2, 35.1)	0.58
	A&E2	-13 (-40.8, 19.5)	0.44
	Total	1 (-19.8, 22.2)	0.91
Specificity	*A&E1*	-3 (-9.7, 3.5)	0.37
	A&E2	-4 (-12.9, 3.7)	0.33
	Total	-3 (-8.8, 1.1)	0.17
Accuracy	*A&E1*	-1 (-10.1, 8.3)	0.84
	A&E2	-1 (-11.8, 9.4)	0.82
	Total	-1 (-7.9, 5.9)	0.77

Table 4.5 shows that before the introduction of CSRs, the GP plain radiograph specificity was high, in the nineties, and sensitivity considerably lower in the fifties. Sensitivity improves after the introduction of CSR reporting and specificity falls. It is interesting to note that whereas the reporting accuracy of A&E1 radiographs was higher than for A&E2 radiographs, the opposite was found for the reporting of GP radiographs. Overall, GP plain radiograph accuracy was almost identical before and after the introduction of the CSR. Indeed, Table 4.6 shows for GP plain radiograph reporting accuracy there was no significant difference (P=0.91).

Table 4.5 GP plain radiograph reporting performance

BEFORE	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
GP1	32	36	25	2	95	56 (43, 68)	95 (83, 99)	72 (62, 80)
GP2	28	56	17	4	105	62 (48, 75)	93 (84, 97)	80 (71, 87)
Total	60	92	42	6	200	58 (49, 68)	94 (87, 97)	76 (70, 81)
AFTER	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
GP1	35	33	21	6	95	63 (49, 74)	85 (70, 93)	72 (62, 80)
GP2	37	46	17	5	105	69 (55, 79)	90 (79, 96)	79 (70, 86)
Total	72	79	38	11	200	65 (56, 74)	88 (79, 93)	76 (69, 81)

Table 4.6 Test for difference in GP plain radiograph reporting performance

Statistic	Body area	% Difference (95% CI)	P-value
Sensitivity	*GP1*	8 (-11.7, 24.1)	0.49
	GP2	7 (-12.4, 24.9)	0.51
	Total	7 (-6.4, 19.5)	0.32
Specificity	*GP1*	-10 (-25.5, 4.2)	0.29
	GP2	-3 (-15.3, 7.7)	0.73
	Total	-6 (-15.3, 2.3)	0.15
Accuracy	*GP1*	0 (-12.8, 12.8)	0.99
	GP2	-1 (12.0, 10.1)	0.86
	Total	0 (-8.9, 7.9)	0.91

4.5.2 Effect on patient management

4.5.2.1 Effect on A&E patient management

In total there were 56 discordant A&E plain radiograph reports, i.e. 27 before and 29 after the introduction of CSR reporting. However, the clinical details for two of the patients before and five after the introduction of CSR reporting were not available, and were excluded from the analyses. A sub-sample of 31 concordant cases was selected to reduce the potential for expectation bias. As it was a convenient sample and not necessarily representative of all the concordant cases, they were excluded from the analyses.

Findings of the effect of incorrect reports on A&E patient management are shown in Tables 4.7 to 4.9 and section 4.4.7.1 defines the three different management options referred to in these tables. It is important to note that, because of the small number of cases, the findings were not very precise and this is reflected in the wide CI. Figures 4.3 to 4.5 illustrate these findings graphically.

Table 4.7 highlights that in reporting of A&E1 radiographs, the percentage of cases falling within the three management options were identical both before and after the introduction of CSR reporting, i.e. zero per cent difference. Moreover, out of a total sample of 201 A&E1 radiographs, there were only two erroneous reports before and after the introduction of CSR reporting that may have resulted in a clinically important difference in patient management. However, the CI is very wide: with 95 per cent certainty, the true population difference in percentage of cases before and after the CSR began reporting that may result in a clinically important difference in patient management, is in the range -34.8 per cent and 34.8 per cent.

Table 4.7 Effect on A&E1 patient management

Management	Before (%)	After (%)	% Difference (95 % CI)
No difference	9/11 (82)	9/11 (82)	0 (-34.8, 34.8)
Clinically unimportant	0/11 (0)	0/11 (0)	0 (-26.8, 26.8)
Clinically important	2/11 (18)	2/11 (18)	0 (-34.8, 34.8)

Figure 4.3 Effect on A&E1 patient management

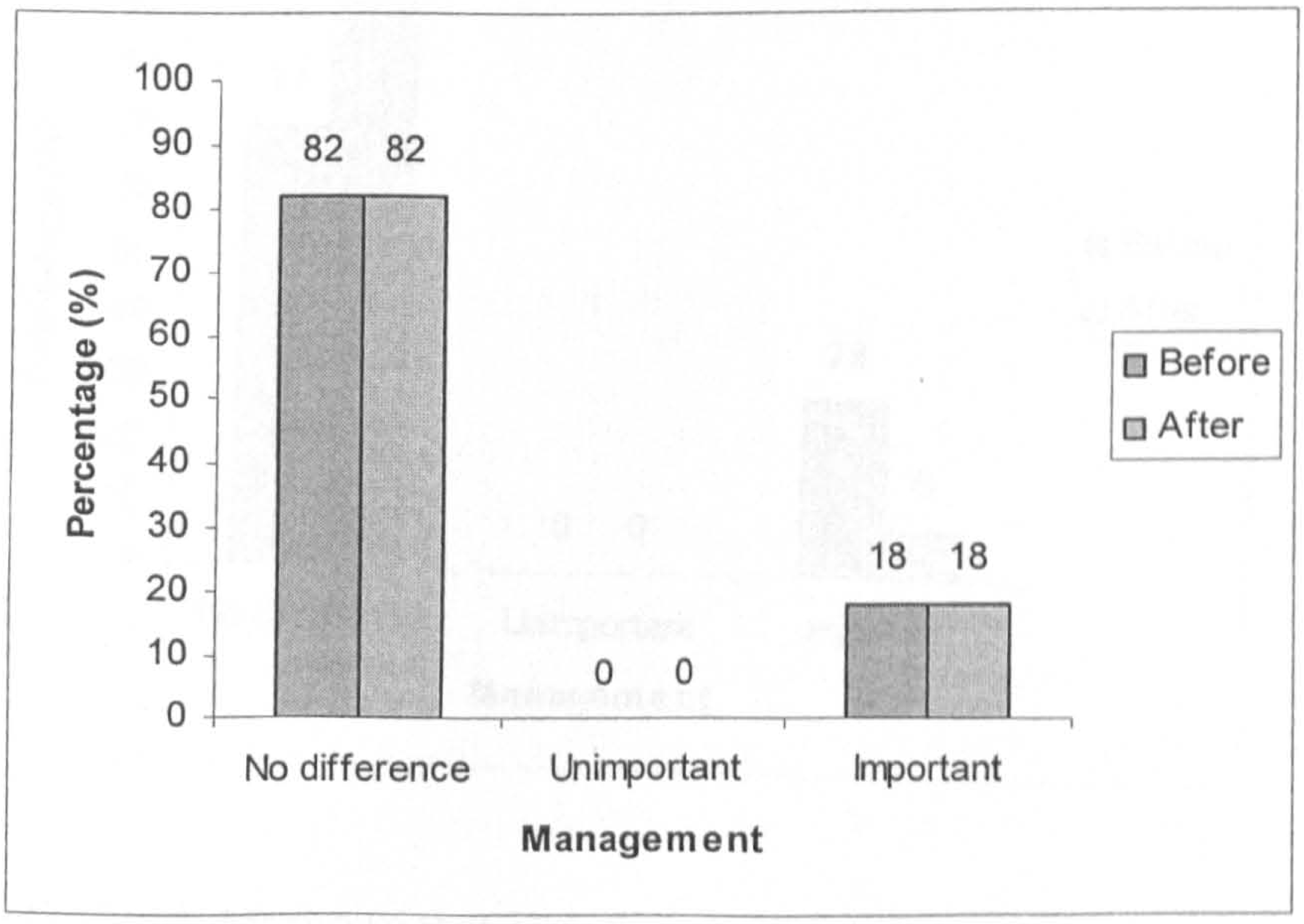
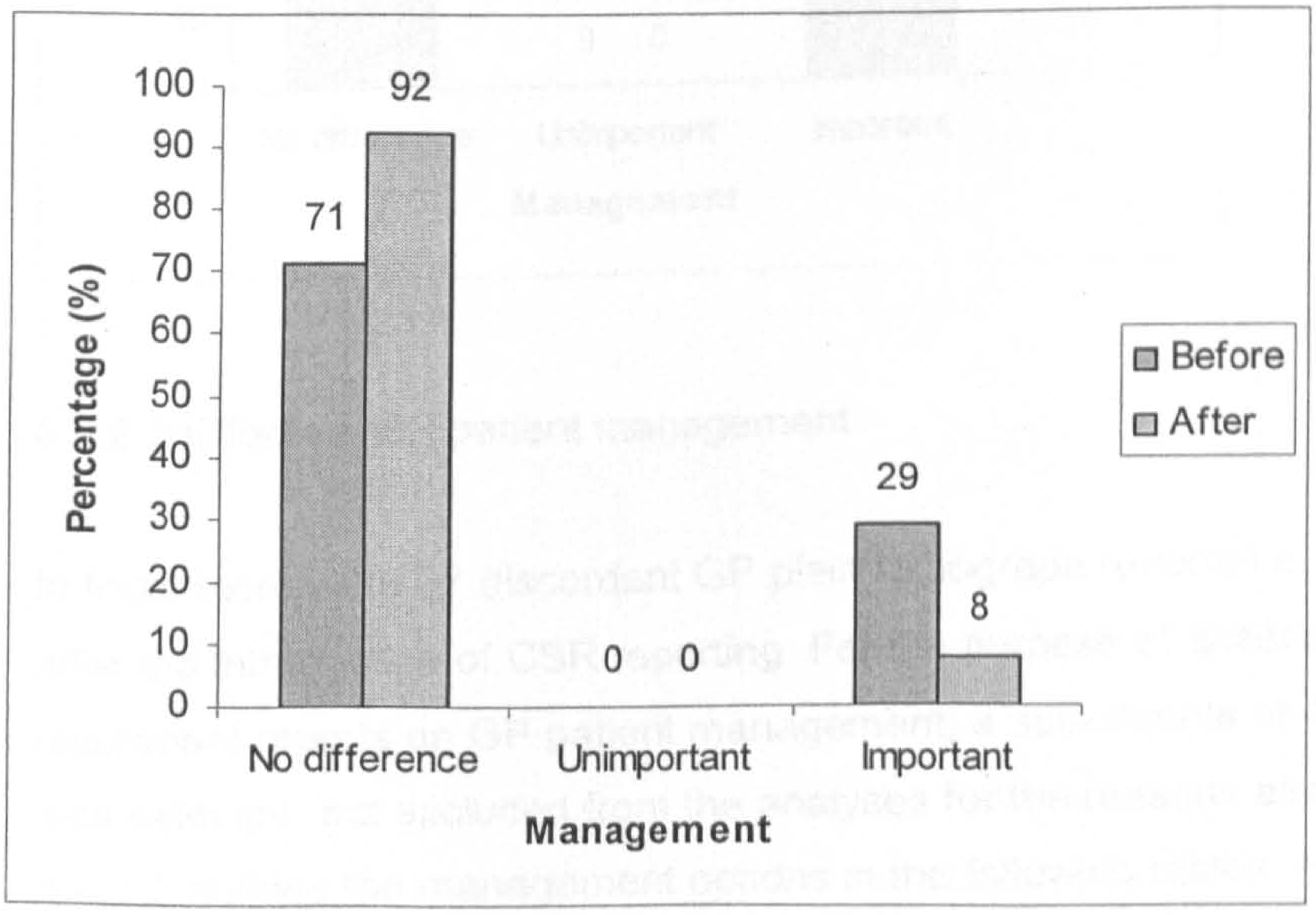


Table 4.8 shows a 21 per cent reduction in A&E2 cases that may have resulted in a clinically important difference in patient management after the introduction of CSR reporting. The CI cross zero, - evidence that there was no significant difference before and after the introduction of CSR reporting, but again, the small sample size is reflected in the wide CI. Indeed, there is 95 per cent certainty that the true population difference in percentage of A&E2 cases before and after the CSR began reporting that may have resulted in a clinically important difference in patient management, was between –49.5 per cent and 10.6 per cent.

Table 4.8 Effect on A&E2 patient management

Management	Before (%)	After (%)	% Difference (95% CI)
No difference	10/14 (71)	12/13 (92)	21 (-10.6, 49.5)
Clinically unimportant	0/14 (0)	0/13 (0)	0 (-22.2, 23.5)
Clinically important	4/14 (29)	1/13 (8)	-21 (-49.5, 10.6)

Figure 4.4 Effect on A&E2 patient management

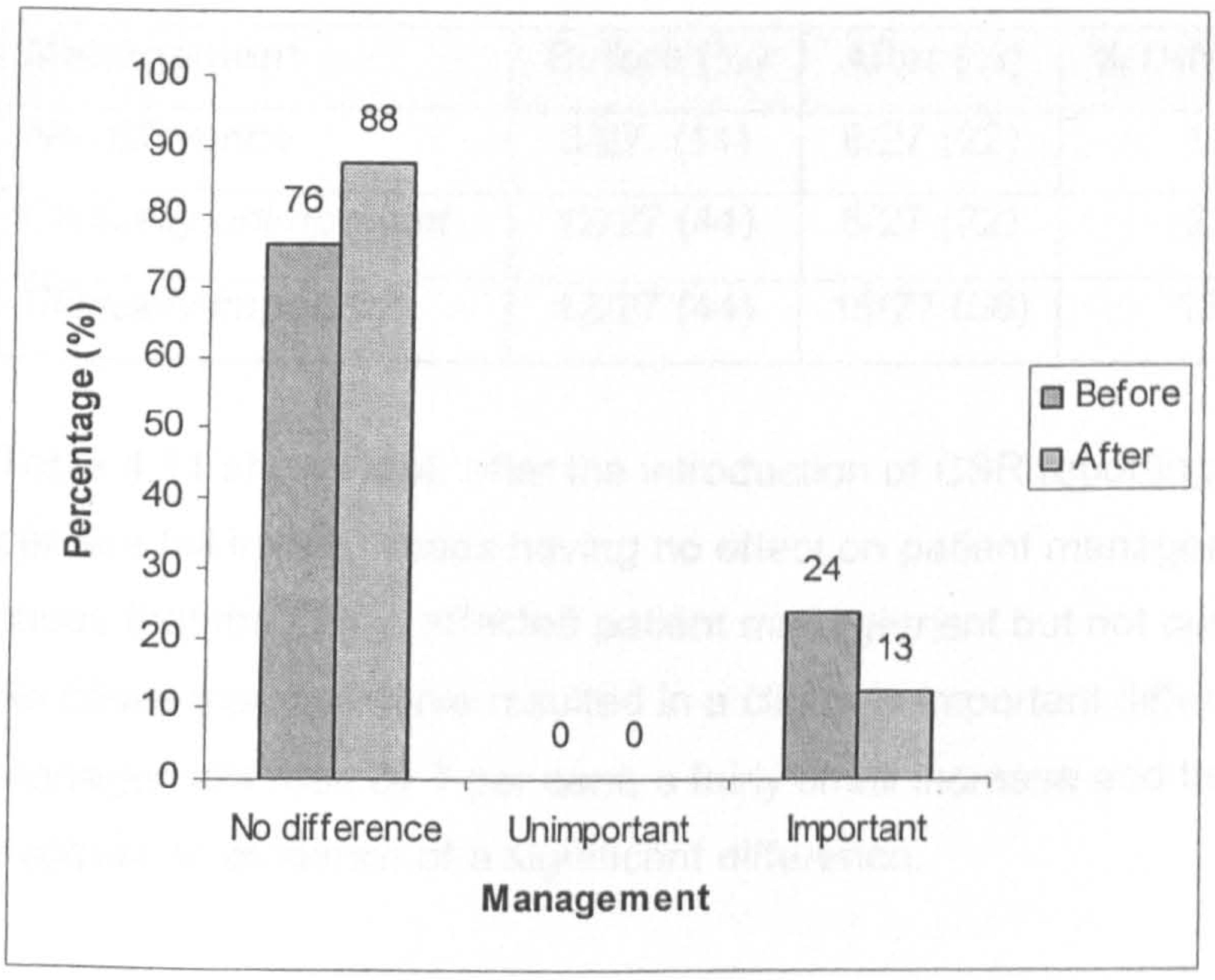


Finally, Table 4.9 shows that after the introduction of CSR reporting, there was an 11 per cent reduction in the percentage of A&E cases that may have resulted in a clinically important difference in patient management. The CI cross zero, which again is evidence that there was no significant difference in A&E patient management before and after the introduction of CSR reporting.

Table 4.9 Effect on A&E patient management

Management	Before (%)	After (%)	% Difference (95 % CI)
No difference	19/25 (76)	21/24 (88)	12 (-11.3, 33.7)
Clinically unimportant	0/25 (0)	0/24 (0)	0 (-13.6, 14.0)
Clinically important	6/25 (24)	3/24 (13)	-11 (-33.7, 11.3)

Figure 4.5 Effect on A&E patient management



4.5.2.2 Effect on GP patient management

In total, there were 97 discordant GP plain radiograph reports i.e. 48 before and 49 after the introduction of CSR reporting. For the purpose of assessing the effect of discrepant reports on GP patient management, a sub-sample of 48 concordant cases was selected, but excluded from the analyses for the reasons already given. Section 4.4.7.2 defines the management options in the following tables. Again, only small numbers are involved so the findings should be interpreted with caution. Furthermore, this only reflects a comparison of the expected management of patients based on the original and reference standard reports, as it was not feasible to collect data on the actual management that patients received by the referring GP.

Table 4.10 shows that after the introduction of CSR reporting, there was an 11 per cent increase in GP1 cases that would have had no effect on patient management. Cases that may have resulted in a clinically important difference in patient management increased by 12 per cent. These increases are explained by the 22 per cent reduction in cases that would have changed patient management but not affected outcome. All the CI cross zero, which suggests that there was no significant difference in the management of GP1 patients before and after the introduction of CSR reporting.

Table 4.10 Effect on GP1 patient management

Management	**Before (%)**	**After (%)**	**% Difference (95% CI)**
No difference	3/27 (11)	6/27 (22)	11 (-9.8, 32.0)
Clinically unimportant	12/27 (44)	6/27 (22)	-22 (-45.2, 3.2)
Clinically important	12/27 (44)	15/27 (56)	12 (-15.5, 36.2)

Table 4.11 shows that, after the introduction of CSR reporting, there was only a one percent fall in GP2 cases having no effect on patient management at all. In addition, cases that may have affected patient management but not outcome fell by 6 per cent. So cases that may have resulted in a clinically important difference in patient management rose by 7 per cent, a fairly small increase and the CI, despite being wide, suggest no evidence of a significant difference.

Table 4.11 Effect on GP2 patient management

Management	**Before (%)**	**After (%)**	**% Difference (95% CI)**
No difference	7/21 (33)	7/22 (32)	-1 (-29.2, 26.2)
Clinically unimportant	5/21 (24)	4/22 (18)	-6 (-30.8, 19.6)
Clinically important	9/21 (43)	11/22 (50)	7 (-22.3, 35.4)

Finally, Table 4.12 shows that for GP management overall, the main change was a 15 per cent reduction in cases that would have changed patient management but not affected outcome. This was explained by a 5 per cent increase in the percentage of cases that would not have affected patient management at all and a 9 per cent increase in cases that may have resulted in a clinically important difference in patient management. For the latter there is 95 per cent certainty that the true population difference is in the range –10.6 per cent and 28.5 per cent.

Table 4.12 Effect on GP patient management

Management	Before (%)	After (%)	% Difference (95 % CI)
No difference	10/48 (21)	13/49 (26)	5 (-11.5, 22.7)
Clinically unimportant	17/48 (35)	10/49 (20)	-15 (-32.4, 3.0)
Clinically important	21/48 (44)	26/49 (53)	9 (-10.6, 28.5)

4.5.3 Effect on A&E patient outcome

As already discussed, 56 erroneous A&E plain radiograph reports were identified - 27 before and 29 after introducing CSR reporting. The sub-sample of concordant cases were included in the method for assessing the effect on patient outcome but removed from the final analyses for reasons already given. No patients re-attended the A&E department within three months of the original consultation.

The following discussion concerns the possible outcome of a reporting error in terms of patient re-attendance to the X-ray department. Table 4.13 describes these outcomes and the number and percentage of cases in which each outcome occurred are presented in Table 4.14.

Table 4.13 Types of patient outcome due to the original report

Outcome
i No further examination (i.e. a patient did not re-attend)
ii No further relevant examination (i.e. re-attend for an unrelated examination)
iii Re-attended for related examination and subsequent report concordant with the original report
iv Re-attended for related examination and subsequent report discordant with the original report

Table 4.14 Patient outcome due to the original report

Outcome	No. (%)	Outcome	No. (%)	% Difference
Before (n=27)		*After (n=29)*		
i	14/27 (52)	i	11/29 (38)	-14
ii	8/27 (30)	ii	13/29 (45)	15
iii	1/27 (4)	iii	1/29 (4)	0
iv	4/27 (15)	iv	4/29 (14)	-1

Notably, Table 4.14 shows that the consultant radiologist at Trust A judged a report subsequent to the original report of an A&E radiograph to be discordant for only 4 out of 27 cases (15 per cent) before and 4 out of 29 cases (14 per cent) after CSR reporting was introduced. A consensus review was conducted so that the three consultant radiologists at Trust A could confirm whether the original report for these cases had been erroneous. Table 4.15 presents the three possible outcomes of the consensus review and Table 4.16 presents the number of cases in which each outcome occurred.

Table 4.15 Types of patient outcome of consensus review

Outcome	
a	Re-attended for related examination but radiographs unavailable e.g. missing
b	Following consensus review no error was found in the original report
c	Following consensus review an error was found in the original report e.g. occult fracture diagnosed on the second visit, fracture missed on the initial visit

Table 4.16 Patient outcome of consensus review

Outcome	No.	Outcome	No.
Before (n=4)		*After (n=4)*	
a	3/4	a	0/4
b	0/4	b	3/4
c	1/4	c	1/4

Table 4.16 shows four cases before the introduction of CSR when patients re-attended and the follow-up report was discordant with the original. In three cases, radiographs were not available to investigate whether the patients' re-attendance was as a result of an error in the first report. One case saw the original report judged discordant. After the CSR were introduced, there were three cases when the follow-up reports agreed with the original, and one case when the original report was judged discordant.

Before the CSR were introduced, the one discordant report was for Case 210, an elderly female patient who had fallen on her hip and had a pelvis plain radiograph. The original report stated 'Degenerative changes are seen at the right hip. No fracture seen'. After ten days, the patient had a further plain radiograph of the pelvis. The follow-up report stated 'There is an impacted fracture of the neck of the right femur'. The consensus review panel found that on both the original and follow-up radiographs 'There is a subcapital fracture'.

The case after the CSRs were introduced was 674, a young male patient, who had both a skull and facial bones radiograph. The original report stated 'No fracture' but the patient had a follow up plain radiograph of the facial bones two days later, in which the report stated 'There is a fracture through the ramus on the left with an associated 2^{nd} fracture more superiorly on the right'. The consensus review found a 'bilateral mandibular fracture' on both the original and follow-up radiographs.

In summary, there were two cases when an overt fracture was missed on the first attendance and then diagnosed at the second visit (i.e. false negative). For both cases, it was a radiologist who had made the original false negative report.

4.6 Discussion

The study aimed to assess the effect of introducing selectively trained radiographers on A&E radiograph reporting accuracy, and the subsequent effects on patient management and outcome. Findings presented in this chapter show that following the introduction of the CSRs, there was insufficient evidence to reject the null hypothesis that there is no statistically and clinically important change in the accuracy with which A&E plain radiographs were reported at Trust A. Furthermore, the analyses of the effect of incorrect reports of A&E plain radiographs on patient management provided no evidence to suggest a significant difference before and after the introduction of CSR reporting. In terms of outcome, none of the patients judged to have an incorrect report re-attended the A&E department within three months of the original examination.

Within a year of the initial radiograph, only one A&E patient before and one after the introduction of CSR reporting had a fracture missed resulting in their re-attending the X-ray Department. In both cases, a radiologist was responsible for the incorrect report.

In addition, a sample of GP reports was used as the non-equivalent, non-intervention control group, which also gave no evidence of a significant difference in GP radiograph reporting performance before and after the introduction of CSR reporting. Nor was there evidence to suggest a significant difference in the effect of incorrect reports of GP plain radiographs on patient management before and after the introduction of CSR reporting. These findings helped to rule out historical threats to study validity, such as a change in radiologists, or inconsistency in arbiter decision-making, or variation in case-mix of radiographs before and after the introduction of CSR reporting, explaining the results of the study.

Some other findings were also noteworthy. First, the systematic review presented in Chapter 2 found that selectively trained radiographers accurately reported A&E plain radiographs during clinical practice at greater than 90% accuracy. But this study found that the *accuracy for A&E and GP radiograph reporting* was in the high eighties and seventies respectively, and explained for the most part by low sensitivities around 50 per cent: one in two abnormal radiographs were missed (i.e. undercalling).

In the A&E reporting system employed at Trust A, a CSR or radiologist only reports radiographs which an A&E clinician has interpreted as normal or insignificantly abnormal. There are no or very subtle abnormalities present on these radiographs, which could explain the high specificity and low sensitivity, but the low sensitivity for GP plain radiograph reporting could not be explained in the same way. Therefore, it was more likely that the criteria used here to judge reports for concordance differed from other studies. Chapter 3 described the strict criteria used to judge reports' agreement on the presence, location and type of abnormality, because unlike other studies the onus was on the A&E Registrar or GP to judge whether discrepant reports would affect patient management. Any indication that an abnormality was missed would result in a false-negative report and subsequently in low sensitivity. Consequently, the lower A&E radiograph reporting performance found in this study should not be cause for concern. This explanation for low accuracy is underpinned by the findings discussed in Chapter 2, that variation in diagnostic threshold between studies may explain variation in the results of the studies when attempting meta-analysis.

Also of interest, was the lower reporting performance of GP plain radiographs compared with A&E plain radiographs (Tables 4.3 and 4.5). Furthermore, as shown in Tables 4.9 and 4.12 respectively, while around 20 per cent of incorrect reports of A&E radiographs may have resulted in a clinically important effect on patient management, around 50 per cent of incorrect reports of GP radiographs would have had this result. This is not surprising as simple binary decisions about the presence or absence of a fracture on A&E radiographs are less appropriate when judging the more complex appearances on GP radiographs. Reporting of GP radiographs has more potential for variation between reference standard and observers' reports and subsequently more discordance, as reflected in the lower performance. Nevertheless, a figure of 50 percent of incorrect GP radiograph reports resulting in an adverse affect on patient outcome is quite disturbing, warranting further investigation in a study that assesses the actual rather than expected effect on GP patient management.

The difference in reporting performance between the appendicular skeleton, (the body areas reported by CSR), and remaining body areas for other types of patient merits attention. The sample was stratified by body area. Previous studies have shown that this affects reporting accuracy (Renwick et al, 1991; Robinson et al, 1999), making it important to include a similar ratio of reports for the different body areas in the samples taken before and after the introduction of CSR reporting. Table 4.3 shows that A&E radiographs of the appendicular skeleton, or A&E1, were reported more accurately than the remaining body areas, A&E2. In contrast, Table 4.5 shows that GP radiographs of the appendicular skeleton, or GP1, were reported less accurately than the remaining body areas, GP2. A possible explanation for this might be that almost 50 per cent of GP2 radiograph referrals were of the spine, many of which would be querying suspicion of osteoarthritis. This condition is known to be present in the elderly and is positively correlated with age, so difficult to incorrectly diagnose. This in turn may explain the substantially higher sensitivity when professionals interpreted GP2 radiographs compared with A&E2 radiographs. The sensitivity with which the appendicular skeleton radiographs were reported was almost identical for both A&E and GP referrals.

The study has its limitations, the first of which is low power, as a result of the small sample size. Results should be interpreted with caution, to avoid incorrect declaration of the presence of 'significant difference'. There was also potential for selection bias. The same person, namely the author, applied the eligibility criteria and randomly selected the reports and radiographs for inclusion in the study. The author could, had he wished, have ignored the random allocation and with access to reports, tried to select a less difficult sample of radiographs after the CSR were introduced. However, the process of assessing eligibility and implementing the allocation had to be undertaken, and with limited resources no other method was feasible. Also, the radiologist who generated the reference standard report had expertise in skeletal radiology. Arguably, other expertise may have been more appropriate for reporting on chest and abdomen radiographs, but other studies have acknowledged that a single consultant radiologist is an acceptable reference standard (Loughran, 1994; Brealey et al, 2002a). Furthermore, evidence in Chapter 3 and the opinion of the other radiologists at the X-ray Department was that an acceptable reference standard was used.

Other limitations include the fact that only the hypothetical effect on GP patient management was assessed, rather than actual management. The retrospective collection of data from A&E patient case notes might affect the completeness and accuracy of information recorded in patient case notes and may therefore restrict the extent to which a relationship between the report and outcome of the patient can be demonstrated. There were also notes missing for two A&E patients before and five A&E patients after the introduction of CSR reporting, but it is unlikely that the data missing for these few cases would affect the results. The method of assessing patient outcome was also limited with the loss of three radiographs before CSR reporting was introduced, preventing the consensus review. Furthermore, patient re-attendance as an adverse event and proxy for patient outcome ignores false positive reports and the morbidity of patients who suffer but do not re-attend. A patient may also re-attend for several reasons other than the original missed abnormality, such as re-injury or a new injury of the affected area, failing to adhere to the original management or receiving inappropriate management from the outset. Many confounding factors could explain why the patient re-attended other than an initial incorrect report, but this clinical follow-up of patients was the only method available to assess patient outcome. It was not possible, for example, to measure change in patient quality of life in a retrospective study. The generalisability of the study was also limited as it only reflected the results of introducing two selectively trained radiographers to an A&E rota in a district general hospital setting. The findings should be interpreted in this context.

Despite these limitations, this study does make a valuable contribution to the evidence base. It was more rigorous in design than its predecessors. As discussed in Chapter 2, previous studies of radiographer plain radiograph reporting are susceptible to bias and do not adhere to methodological standards espoused by papers on the evaluation of diagnostic tests. An attempt was made to eliminate all sources of bias and adhere to all relevant methodological standards whilst maintaining a pragmatic design. The study was also enhanced by the method of selecting reports such that the sample was representative of the population in clinical practice and comparable both before and after the CSR began reporting. Further underpinning was provided by an additional control group, which helped to eliminate potential threats to study validity. Despite the low power, only a one per cent fall in A&E radiograph reporting accuracy was detected before and after the introduction of CSR reporting - a finding neither statistically or clinically significant. Evidence is also provided that the introduction of CSR reporting had no obvious detrimental effect on A&E patient management and outcome. Taken together, these findings provide evidence of the successful introduction of CSR reporting A&E plain radiographs of the appendicular skeleton at Trust A.

4.7 Conclusion

There is no statistically or clinically significant difference in A&E radiograph reporting performance following the introduction of the CSRs. Neither was there evidence to suggest a statistically significant difference in A&E patient management, nor obvious difference in patient outcome in terms of patient re-attendance to A&E or the X-ray Department. Despite the small sample size and the other limitations described, this study does provide evidence that the CSRs were successfully introduced to the reporting service. The implications for clinical practice and policy are reserved until the findings from the other studies are discussed in Chapter 7.

Finally, having assessed the introduction of CSR to the A&E reporting service at Trust A, it was deemed appropriate to assess the potential for extending their role to include the A&E plain radiographs for the remaining body areas and all GP plain radiographs, as presented in the next chapter.

Chapter 5

Radiographers and Radiologists reporting A&E and GP plain radiographs: A quasi-experimental study

5.1 Introduction

Chapter 2 reviewed the existing evidence about radiographer plain radiograph reporting performance, with a discussion in Chapter 3 of the methods used to measure reporting performance for the primary studies in Chapters 4 and 5. Chapter 4 described a pragmatic study on the effect of introducing clinical specialist radiographers (CSRs) reporting accident and emergency (A&E) appendicular skeleton plain radiographs at Trust A. The aim of the study presented in Chapter 5 was to assess the potential for extending the CSRs reporting role to include the body areas for A&E patients they do not currently report and also for plain radiographs referred by general practitioners (GPs). The rationale for this study follows.

When the study was conceived, the CSRs at Trust A had been reporting A&E plain radiographs of the appendicular skeleton for two years, so staff in the X-ray department wanted to determine the potential for further extending their reporting role without additional training. At the time, there was no evidence of selectively trained radiographers reporting GP plain radiographs, so it was agreed to assess their ability to report these radiographs accurately. This decision was justified by the results of the systematic review presented in Chapter 2, which showed that no study has assessed selectively trained radiographers reporting solely on GP radiographs. Such a study was also desirable in light of the College of Radiographers' (1997) vision that all radiological examinations carried out by radiographers should receive a radiographer report.

As with the study presented in Chapter 4, the absence of evidence from the systematic review regarding the clinical effects of plain radiograph reporting made this an important objective to address here. Chapter 4, though, presented a retrospective design, so data from clinical practice was used to assess the effect of CSR reporting on patient management and outcome. In contrast, *this* study was not conducted during clinical practice, enabling prospective data collection, and allowing a more detailed, but hypothetical investigation into the clinical effects of CSRs and radiologists reporting plain radiographs. It included not only an assessment of reporting accuracy but the effect of CSR and consultant radiologist reports on clinicians' choice of diagnosis and confidence therein, confidence in choice of patient management, and effect on patient outcome. It was beyond the scope of this study to assess the associated cost of CSRs reporting GP plain radiographs.

The *primary* objective was to assess the performance of selectively trained radiographers, or CSRs, and consultant radiologists in reporting A&E and GP plain radiographs. The *secondary* objective was to assess the subsequent clinical effects of their incorrect reports on clinicians' choice of diagnosis and confidence therein, confidence in patient management, and effect on patient outcome. Results should contribute to the evidence about selectively trained radiographers reporting A&E plain radiographs and the continuing debate on the extent to which their reporting role should be extended to substitute or complement other health care professionals.

Finally, in contrast to the pragmatic study presented in Chapter 4, this was an explanatory study designed to ascertain whether CSR reporting of A&E and GP plain radiographs was efficacious. Currently, CSRs do not report most of these radiographs in clinical practice, making it important to establish to begin with, whether they could report these radiographs under controlled conditions when there would be no actual effect on patient management and outcome. Results could then inform a decision on the need for a future study to assess their performance during clinical practice. This was therefore a feasibility study to a future, larger, pragmatic study assessing the cost-effectiveness of CSR reporting these plain radiographs.

5.2 Objectives

5.2.1 Primary objective

1. To assess CSR and consultant radiologist performance when reporting A&E and GP plain radiographs compared with a reference standard.

5.2.2 Secondary objective

2. To assess the effect of CSR and consultant radiologist incorrect reports on clinicians' diagnosis and confidence therein, confidence in patient management, and patient outcome.

5.3 Framework for assessing the effects of CSR and radiologist reporting

The first four chapters of this thesis have, to a varying extent, contained discussion of the complexities of defining a report and the problems with measuring reporting performance. For the purpose of this study, which was to assess in more detail the clinical effects of CSR and radiologist reporting, the conceptual hurdle to be overcome was how to relate a report to patient outcome when factors such as therapy are involved. The purpose of this section was to describe the framework used to assess the plain radiograph reporting accuracy of the CSRs and consultant radiologists and the subsequent clinical effects, before discussing the specific methods used to address the study objectives in the next section.

It was appropriate in considering the effect of a report on patient outcome to refer to the framework proposed by Fineberg et al (1977) to address the debate about the adoption of Computed Tomography in the 1970s. To measure the chain of events between the application of a diagnostic technology and any potential influence on the disease, they defined this concept at four separate levels. The framework was subsequently extended to five levels by the Institute of Medicine (1977) and more recently to six by Fryback & Thornbury (1991). This framework has been applied to the assessment of radiographer reporting as illustrated in Figure 5.1 (Brealey, 2001a).

The first level of the framework presented in Figure 5.1 is concerned with the 'Technical Competence' of health care professionals and their *potential* for reporting radiographs. Eye-tracking equipment is used in a laboratory controlled setting to monitor visual search patterns of different professional groups. Carr & Mugglestone (1997) recorded the visual search behaviour of radiographers when viewing chest radiographs in experimental conditions. They found that radiographers had comparable patterns of search strategies to radiologists and achieved a high rate of agreement about the presence or absence of abnormalities. There is now evidence that selectively trained radiographers can accurately report in clinical practice and at no significant difference from radiologists (Robinson, 1996a). Radiographers, then, have been demonstrated to show similar visual search behaviour to radiologists when identifying abnormal appearances. Training then provides the clinical knowledge, skills and experience to enable them to interpret those appearances. This study did not investigate the 'Technical Competence' of the CSRs, as they had already received training in image interpretation and had experience of reporting plain radiographs in clinical practice. The final level of the hierarchy, or 'Societal Level', goes beyond the clinical effect of different health care professionals reporting to consider whether the cost is acceptable to society. The extent to which reporting is an efficient use of resources would depend not only on the direct costs to the NHS, but also the personal costs borne by patients and their families, such as time and travelling expenses. The extent to which CSRs reporting A&E plain radiographs is the most prudent or optimal mix of resources is explored in Chapter 6.

Figure 5.1 *The evaluative hierarchy as applied to assessing radiographer reporting*

Technical Competence

Do radiographers use visual search patterns comparable with that of an expert?

↓

Diagnostic Performance

Do radiographers accurately interpret radiographs compared with a reference standard?

Do radiographers consistently agree with the expert observers in clinical practice?

↓

Diagnostic Outcome

Does radiographer reporting improve clinician's diagnostic confidence and understanding?

↓

Therapeutic Outcome

Does radiographer reporting contribute to the planning and delivery of therapy?

↓

Patient Outcome

Does radiographer reporting improve patient health?

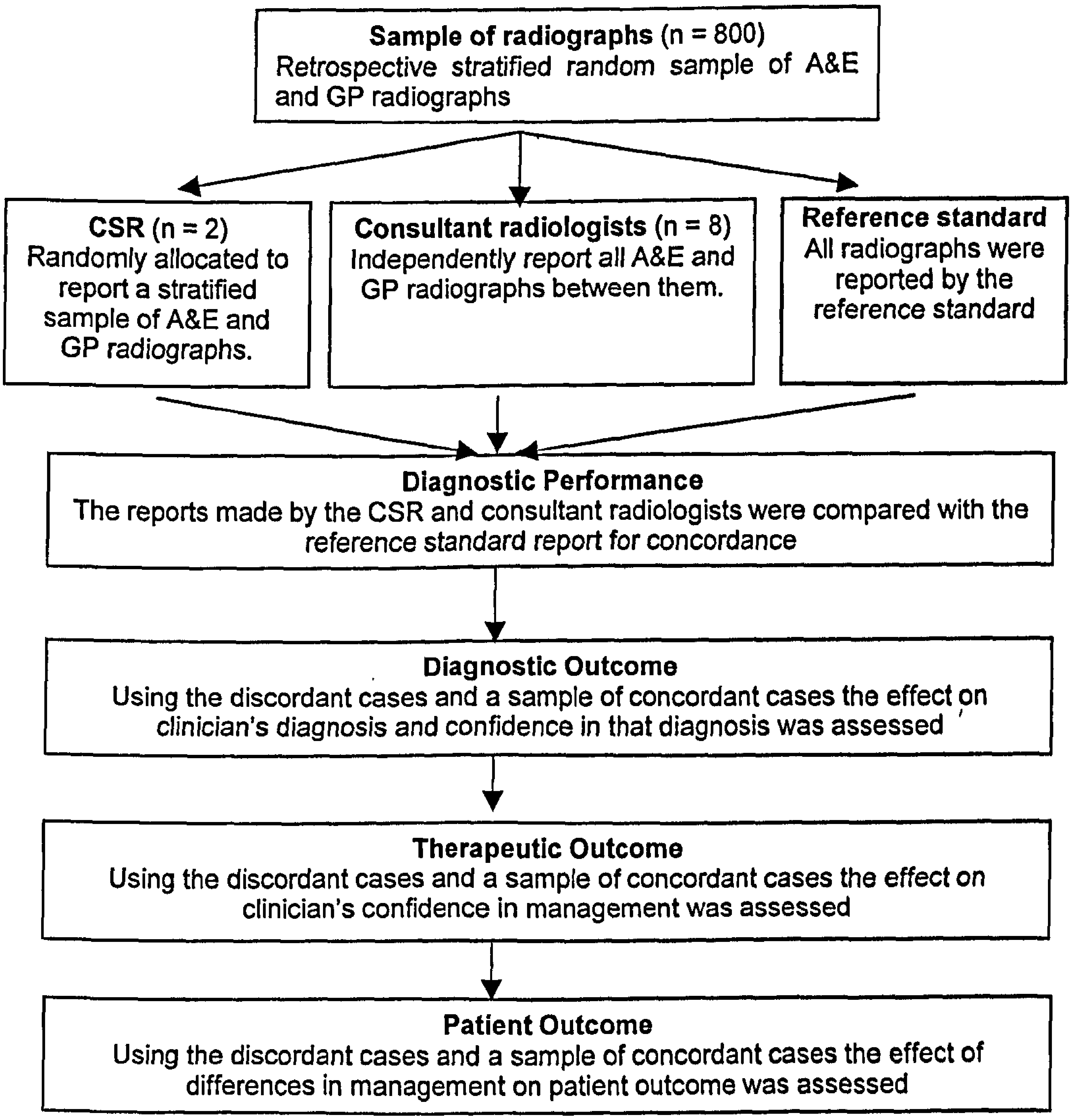
↓

Societal Level

Is the cost (borne by society as a whole) of radiographer reporting acceptable?

To address the objectives of this chapter, the levels of the framework used to assess the two professional groups were 'Diagnostic Performance', 'Diagnostic Outcome', 'Therapeutic Outcome', and 'Patient Outcome'. Diagnostic Performance is concerned with the accuracy of CSRs' and radiologists' reporting of radiographs compared with a reference standard. Diagnostic Outcome and Therapeutic Outcome ask whether the report leads to a change in clinician behaviour. Assessing Diagnostic Outcome may involve recording changes in a clinician's diagnosis and their confidence in that diagnosis based on a report by a CSR or radiologist. Similarly, Therapeutic Outcome may be assessed as a change in the referring clinician's choice in patient management and confidence. Finally, Patient Outcome considers the effect of a report on patient's health, measured in terms of patient quality of life, number of days off work, or the clinician's subjective opinion as to whether a difference between reports could affect patient outcome. Figure 5.2 summarises the plan of investigation for this study.

Figure 5.2 Plan of investigation



5.4 Assessment of diagnostic performance

This section looked at the methods used to address the primary objective of this study about CSRs' and consultant radiologists' performance when reporting A&E and GP plain radiographs compared with a reference standard. It describes the choice of study design to assess the CSR and radiologist reporting performance and the calculation of the sample size required to address the primary objective of the study. The remainder of this section discusses eligibility criteria and methods used to select the radiographs to be reported, the way in which they were to be reported, the use of the reference standard and how the reports were to be compared.

5.4.1 Study design

In the choice between alternative health care policies such as CSR or radiologist reporting, there are essentially three approaches. Observational studies are undertaken when the policies to be compared are observed without intervention. A quasi-experimental approach may be adopted when a decision-maker replaces one policy to be evaluated with another policy and the two can then be compared *as if* the resulting data arose from a scientific experiment. This was the approach used in Chapter 4 to assess the introduction of CSR reporting using a controlled before and after design. Experimental studies require an intervention in the status quo for the sole purpose of evaluating alternative policies and the resulting comparison must exhibit all the essential attributes of a scientific experiment, in particular random allocation to guard against bias and to provide a basis for statistical analysis (Pocock, 1983).

An experimental study at Trust A would require intervention in normal clinical practice to evaluate the CSRs compared with consultant radiologists for body areas and patients they do not currently report. This would raise both ethical and practical objections, for example, it would not be acceptable for CSRs to report GP referred plain radiographs without further training. The design would also require considerable re-organisation within the X-ray department. This was unfeasible without financial support and not desired by the department. However, a retrospective sample of plain radiographs was readily available for reporting without the need for patient consent, allowing the principles that underpin a randomised controlled trial (RCT) to be applied so that the two professional groups could be compared *as if* the resulting data arose from an experimental study. Hence, this study is described as quasi-experimental in design.

5.4.2 *What sample size was necessary?*

Table 5.1 describes the different groups of radiographs reported by the CSRs and radiologists. The same sample of radiographs was included as in the previous study, so the sample size was fixed.

Table 5.1 Description of the different sample of radiographs

Group	Source	Number	Body Areas
A&E1	A&E	201	Upper limb, lower limb
A&E2	A&E	199	Axial skeleton, chest, thorax, abdomen
GP1	GP	190	Upper limb, lower limb
GP2	GP	210	Axial skeleton, chest, thorax, abdomen

To address the primary objective of the study, the main analysis of interest was testing for a significant difference in the performance of the CSRs and consultant radiologists when reporting the different groups of radiographs. When discussing the calculation for the study presented in Chapter 4, it was demonstrated that such an analysis would require several hundred radiographs in each group rather than two hundred.

It is essential to note that this was a feasibility study, assessing the potential for extending selectively trained radiographers' reporting role in controlled conditions. The results could then be used to inform the need for a larger study in the future concerning cost-effectiveness of CSRs' reporting these radiographs during clinical practice. Consequently, a tentative attempt was made to calculate the sample size required for the secondary analysis of interest, which was to test for a significant difference between CSRs' reporting of A&E radiographs that they do (A&E1) and do not (A&E2) currently report. This was the next most appropriate comparison to make, as the initial natural extension of the CSR reporting role at Trust A would be to report the remaining body areas for A&E radiographs.

Calculation of the required sample size used evidence from the study by Loughran (1994) who found that during a training period radiographers' detection of fractures improved from 81 per cent to 94 per cent accuracy when reporting A&E radiographs. The assumption seemed reasonable, that if the CSRs at Trust A were already trained in image interpretation and had experience of reporting A&E1 radiographs, they would report 94 per cent of these correctly and around 88 per cent of A&E2 radiographs correctly. Using nQuery 4.0 to have 80 per cent power to detect a 6 per cent difference in the accuracy with which A&E1 and A&E2 radiographs were reported as statistically significant at the 5 per cent level required a sample size of 356 radiographs in each group (or 712 in total). However, the sample size was fixed at 200 radiographs for each group, so the power of the study was only 55 per cent, as shown in Table 5.2.

Low power in this study indicates that results should be interpreted with caution. Nevertheless it was still valuable to conduct for the reasons already given in the introduction to this chapter and because it was a feasibility study performed in controlled conditions the results of which should inform the need for future investment in a larger pragmatic study.

Table 5.2 Sample size calculation

	Required	Actual
Test significance level, α (two-sided)	0.050	0.050
A&E1 proportion	0.940	0.940
A&E2 proportion	0.880	0.880
Power	80	55
N per group	356	198

5.4.3 Eligibility criteria

The same sample of plain radiographs was used as for Chapter 4, so the same eligibility criteria applied.

5.4.4 Selection of radiographs

As the sample of radiographs described in Chapter 4 had already been selected and there was a reference standard report these were used for this study. Furthermore, as a random sample of radiographs had been selected for each group they were a valid representation of the case mix of radiographs in clinical practice and provided a basis for statistical analysis.

The different samples are described in Table 5.1. CSRs currently report A&E radiographs of the appendicular skeleton (A&E1) in clinical practice. They do not report the remaining body areas for A&E patients (A&E2) and GP patients (GP1, GP2). A randomised controlled trial would describe the A&E1 radiographs as the control group and the other radiographs as the experimental groups. But the sample of radiographs the CSRs currently report was not controlling for potential confounding factors, which is the purpose of controls in a RCT. A&E1 radiographs are therefore more like a usual practice group.

Since the study was assessing the potential for extending the CSRs' reporting role, comparison of the independent performance of the two CSRs when reporting the different groups of radiographs was essential. Were only the overall performance of the two professionals presented, it may have masked whether one CSR's performance was worse than the other's. This was not as necessary for the assessment of the consultant radiologists, as they already had considerable training and experience of reporting both A&E and GP plain radiographs.

The most valid method of comparing the individual CSR's performances would be to have them report the same sample of radiographs: any difference could then be attributed to the competence of the individual CSR rather than variation in the case mix of radiographs. It was not feasible for each CSR to report eight hundred plain radiographs, so a random sampling method was used to ensure they reported a comparable case mix. There is evidence that different body areas affect radiograph reading performance (Renwick et al, 1991) so block randomisation was used to ensure they reported a comparable and equal number of radiographs for the different body areas. The body areas within each stratum are illustrated in Table 5.3.

Using a table of random numbers each stratum was randomised into blocks of two radiograph examinations (Pocock, 1983 p.74). Letters AB were assigned to digits 0-4 and BA to digits 5-9, representing CSR A and B, so that numbers 0 5 2 7 etc. would produce a list starting AB BA AB BA etc. Table 5.4 is an example of the randomisation list within each stratum.

Table 5.3 Body areas within each stratum

Strata	Body area
Upper limb	Bones: finger(s); thumb; hand; humerus; radius and ulna; scaphoid; clavicle Joints: acromio-clavicular joint; elbow; shoulder; wrist
Lower limb	Bones: femur; calcaneus; foot; patella; tibia and fibula; toe(s) Joints: ankle; knee; hips
Pelvis	Pelvis
Spine	Cervical, thoracic and lumbar spine; coccyx; sacrum and coccyx; sacrum; sacro-iliac joints
Head	Orbits; facial bones; internal-auditory-meatus; mandible; mastoids; nasal bones; sinuses; skull; tempero-mandibular joint
Chest/Thorax/ Abdomen	Chest; ribs; sternum; thoracic inlet; sterno-clavicular joint; larynx; soft tissue view neck; abdomen; kidneys, ureters and bladder

Table 5.4 An example of random permuted blocks within strata

Block	Strata (body areas)	CSR	
I	Upper limb	AB AB	AB BA
II	Lower limb	BA BA	AB AB
III	Pelvis	BA AB	AB AB
IV	Head	AB AB	AB BA
V	Spine	AB BA	BA AB
VI	Chest/Thorax/Abdomen	BA AB	AB AB

5.4.5 Method of reporting by the CSR and radiologists

The sample of radiographs and their bags were stored in the CSR reporting room. A tag was attached to each bag, labelled with one or more reference numbers ranging from one to eight hundred and the CSRs were provided with a list of the radiographs they had been randomly allocated to report. To prevent bias, the author removed all previous reports from patients' radiograph bags and stored them in another room, although previous plain radiographs were available to appropriately assist interpretation. The CSRs and radiologists used Proforma A5.1 to record their reports. All proformas are presented in Annex 5. During clinical practice, the CSRs occasionally seek further advice from a radiologist. For the purpose of this study, this was not permitted, but after recording their report on the proforma they printed on the back in bold letters whether further advice was required (FAR) and the reason. Written text and the codes listed in Table 5.5 were used to describe the radiograph appearances.

Table 5.5 Reporting codes

Codes	Definition
A3	No bony injury identified
A9	No radio-opaque foreign body seen
C5	No relevant abnormality
S1	No fracture identified. If fracture scaphoid is still clinically suspected in 12-14 days, a repeat X-ray is suggested
DCO	Degenerative changes only
NNF	No new fracture
STS	Soft tissue swelling noted, but no bony injury identified
CH1	Normal heart and lungs
CH3	No free subdiaphragmatic gas or focal lung lesion shown

Both groups reported in similar viewing conditions, but the CSRs reported during normal reporting sessions and the consultant radiologists during audit. The radiographs to be reported by individual radiologists could not be randomly selected, as the author did not know until the beginning of the audit session precisely who was available or how much time was allotted. The author could only select a consecutive block of between 20 to 30 radiographs for individual radiologists to report. No explicit bias in selection was present, as neither the radiologist nor the author made a selective choice of radiographs. Those selected were a convenience sample of the stratified random sample.

5.4.6 Method of reporting by the reference standard

The reference standard report for this sample of radiographs had already been generated, as described in Chapters 3 and 4. To prevent incorporation bias, the consultant radiologist providing the reference standard report was not included in the group of radiologists under evaluation (Brealey & Scally, 2001). When the reference standard report agreed with the report of a consultant radiologist from the previous study, this generated a double blind consultant radiologist report and provided an opportunity to analyse the performance of the CSRs and radiologists for these cases.

5.4.7 Method of comparing reports

The same arbiter as for the previous study compared the reports of the CSRs and consultant radiologists with the reference standard using Proforma A5.1. The arbiter was blind to who made the reports and the position of the reference standard was 'randomly' placed in the left or right box. The marking criteria explained in Chapter 3 were used to judge reports for concordance. Additional marking criteria are described in Annex 5.

5.5 Assessment of diagnostic and therapeutic outcome

The Royal College of Radiologists (RCR) and College of Radiographers (CoR) acknowledge radiographers' ability to provide a description of radiograph findings, but a report on the further medical management of the patient can only be provided by appropriately trained medical practitioners, normally radiologists (RCR & CoR, 1998). Differences in the content and certainty with which professionals report could affect the diagnosis and management decisions of the referring clinician, so assessment was made as to whether discordant reports by the CSRs and radiologists would affect clinician's choice of diagnosis, confidence in their diagnosis and confidence in their choice of patient management.

Clinicians were invited to complete a pre-report questionnaire. This was an A&E Specialist Registrar at St James' University Hospital in Leeds and one of three GPs (two Registrars and one qualified). The questionnaire allowed them to record their diagnosis and confidence on a scale of 0 per cent (uncertain) to 100 per cent (certain), and also their proposed management plan and confidence in that plan. The clinician without access to the original questionnaire and blind to who made which report then completed a post-report questionnaire that included the report of the CSR or radiologist. The A&E clinician judged the expected patient management using the options available at Trust A as listed in Table 5.6.

Table 5.6 A&E management options

Option		Example
1.	Admit patient	e.g. compound fracture
2.	Fracture clinic	e.g. closed Colles fracture
3.	Dressing clinic	e.g. abrasions
4.	GP review	e.g. PIPJ volar plate fracture
5.	Refer to a specialist centre	e.g. neuro
6.	TRIN (To Return If Necessary)	e.g. lateral malleolus flake
7.	Discharge	e.g. no follow up/normal
8.	Specialist clinic (medical)	e.g. ENT
9.	Specialist clinic (non-medical)	e.g. physiotherapy
10.	Injury clinic	e.g. review by A&E consultant
11.	Advice (oral/printed sheet)	e.g. soft tissue swelling only

The pre-report and post-report questionnaires (Proformas A5.3 and A5.4 respectively) are modelled on those used by Wittenberg et al (1980). The collection of this data enabled assessment of changes in diagnosis, management plans and confidence. In completing the process, it was important that each clinician did not know that all reports were discordant as this knowledge may have biased their decision-making. For the same reasons as described in the previous chapter, for every two pairs of reports that disagreed a pair of concordant reports was included giving a 2:1 ratio. All decisions by the clinicians were made using the patient clinical details and the reason for the X-ray examination.

5.6 Assessment of patient outcome

A missed X-ray abnormality such as a fracture of the nasal bone or a tiny flake from the lateral malleolus is not clinically important (de Lacey et al, 1980). However, because of the strict decision-making criteria, this should result in a report being judged incorrect, so the clinicians completed Proforma A5.5 to judge whether discrepancies between the professional's report and the reference standard might result in a difference in patient management affecting patient outcome. The proforma included the clinical details, the pair of reports, and the expected patient management for the two reports. The clinicians then compared the choice of management blind to who made which report. To judge if a difference was clinically significant, one of the three options listed in Table 5.7 was chosen.

Table 5.7 Potential effect on patient outcome

Option
No difference in the expected patient management (treatment or advice) and therefore no effect on patient outcome
A clinically unimportant difference to the expected patient management, i.e. the patient would have received different management as a consequence of the two reports but no effect on patient outcome
A clinically important difference to the expected patient management, i.e. the patient would have received different management as a consequence of the two reports and it might affect patient outcome

5.7 Data analyses

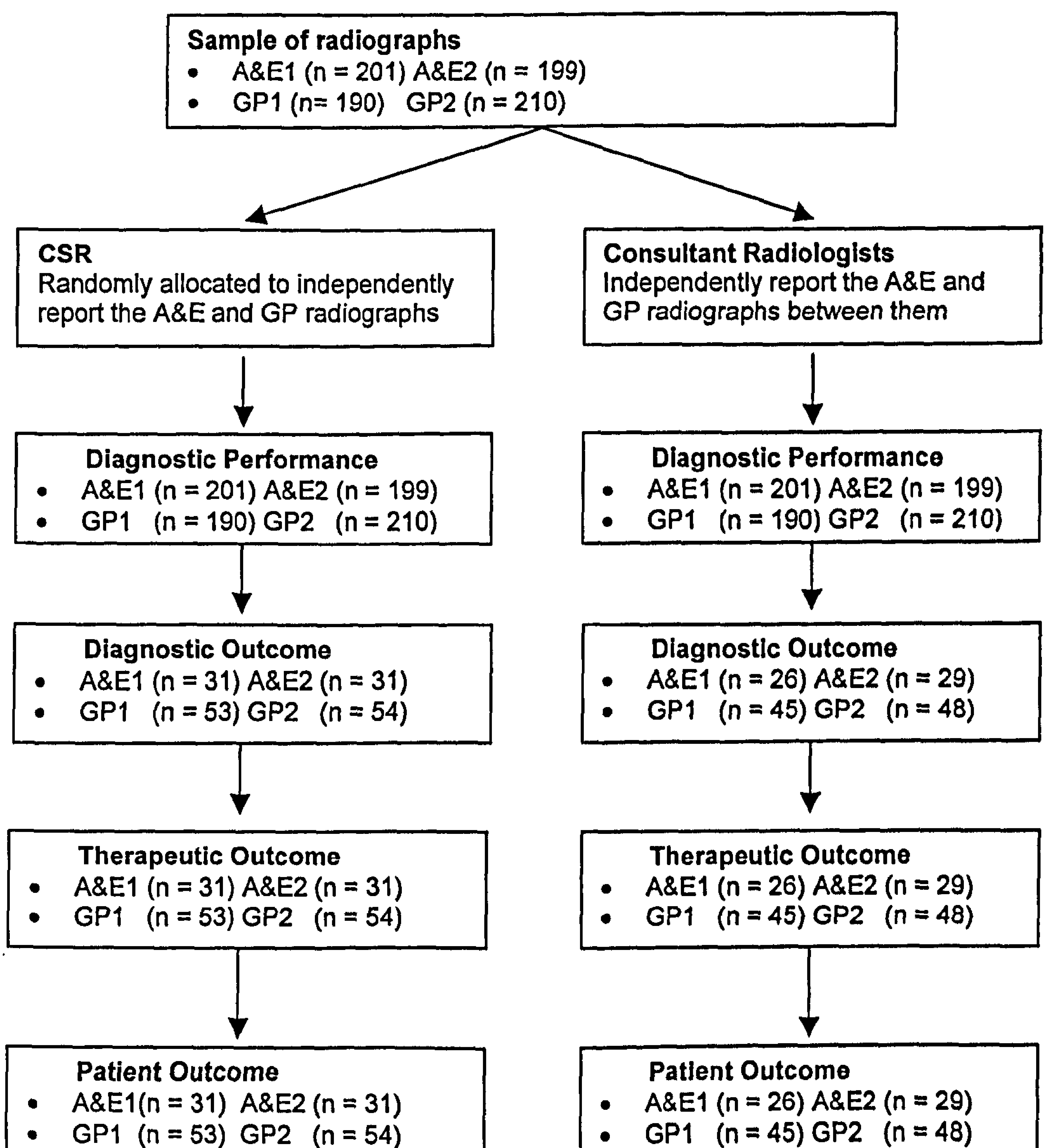
Radiograph reporting performance was calculated using sensitivity, specificity and accuracy and both paired and independent samples of data were included in the analyses. Paired data would include the CSR and radiologist reporting performance being compared for the same sample of radiographs, such as A&E1. Independent data was when the CSR reporting performance was compared for two different samples of radiographs, such as A&E1 and A&E2. Stats Direct was used to calculate the 95 per cent confidence intervals (CIs) around a single proportion and the approximate 95 per cent CIs when testing for a difference in proportions between two independent groups. The formula in Altman (1991) was used to calculate the 95 per cent CIs when testing for a difference in proportions between two paired groups and the appropriate hypothesis tests: McNemar's test for paired data and chi-square test for independent data or Fisher's Exact test if the total number of observations were less than twenty or any of the expected frequencies were less than five. The AccuROC package was also used to produce receiver operating characteristic (ROC) curves and the area under the curve, or Az value. This value is the probability of a plain radiograph being correctly reported as normal or abnormal. An Az of 0.5 indicates guesswork and an Az of 1.0 is perfect (Metz, 1989). A nonparametric method derived from correlated samples was used for comparing the Az values between the CSRs and radiologists (DeLong et al, 1988).

Stats Direct was used in the analysis of the affect of incorrect reports on clinicians' diagnosis and patient outcome to calculate the 95 per cent CI when testing for a difference in proportions between two independent groups. Stats Direct was also used to calculate independent t-tests, to produce the 95 per cent CI around the mean differences in clinicians' confidence in diagnosis and choice of patient management, based on incorrect reports by the two professional groups. The F test was used to test for equal variances between the two independent samples of data and if this proved significant, the approximate t (unequal variances) results were used. All tests were two-sided. The P-values for these tests are not presented, as the small sample sizes could result in erroneous conclusions as to the presence of a significant difference. Instead the emphasis was on interpreting the 95 per cent CIs to indicate precision of the estimated difference in proportions and means.

5.8 Results

Radiograph reporting performance of the two professional groups is presented here, as is the effect of their incorrect reports on the referring clinician's diagnosis and diagnostic confidence (i.e. Diagnostic Outcome) and confidence in their choice of management (i.e. Therapeutic Outcome). Finally, differences in patient management based on the two groups' incorrect reports compared with management based on the reference standard report to assess the effect on Patient Outcome is discussed. Figure 5.3 shows the process of evaluating performance and changes in sample size.

Figure 5.3 Flowchart describing the process of evaluation



5.8.1 *Diagnostic performance*

This section presents the results of addressing the primary objective: assessment of CSR and consultant radiologist performance when reporting A&E and GP plain radiographs compared with a reference standard. The following tables give the performance of the two groups when reporting the radiographs defined in Table 5.1.

5.8.1.1 Comparison with a single consultant radiologist

Tables 5.8 to 5.16 illustrate results based on the single consultant radiologist as the reference standard. These data include all cases, even those when CSRs would have sought further advice from a radiologist if reporting in clinical practice. Table 5.8 shows that both CSRs (A and B) report A&E radiographs to a similar level of accuracy for the body areas they currently report (A&E1) and the areas they do not (A&E2).

Table 5.8 *CSR A&E radiograph reporting performance*

A	Body area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	A&E1	15	70	12	3	100	56 (35, 75)	96 (88, 99)	85 (76, 91)
	A&E2	8	78	8	6	100	50 (25, 75)	93 (85, 97)	86 (78, 92)
	Total	23	148	20	9	200	53 (38, 69)	94 (89, 97)	86 (79, 89)
B	**Body area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	A&E1	15	70	9	7	101	63 (41, 81)	91 (82, 96)	84 (76, 91)
	A&E2	11	71	10	7	99	52 (30, 74)	91 (82, 96)	83 (74, 90)
	Total	26	141	19	14	200	58 (42, 72)	91 (85, 95)	84 (78, 88)

Table 5.9 illustrates that the CSRs' reporting accuracy is comparable to that of the radiologists when reporting A&E radiographs. Indeed, the P-values listed in Table 5.10 show no evidence of a significant difference in accuracy between the two professional groups. In particular, there is no significant difference when reporting the body areas the CSRs do not currently report (P=0.70) and can be 95 per cent confident that the radiologists are between 4 per cent worse and 6 per cent more accurate than the CSRs. But whereas specificity is high and almost identical for both professional groups, the sensitivity for both groups is low, and notably for the CSRs when reporting A&E1 radiographs. The false negative (FN) figures show that the CSRs were 'undercalling'; that is missing more abnormalities than the radiologists.

Table 5.9 *CSR versus radiologist A&E radiograph reporting performance*

CSR	Body Area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	A&E1	30	140	21	10	201	59 (44, 72)	93 (88, 97)	85 (79, 89)
	A&E2	19	149	18	13	199	51 (34, 68)	92 (87, 96)	84 (79, 89)
	Total	49	289	39	23	400	56 (45, 66)	93 (89, 95)	85 (81, 88)
Rad	**Body Area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	A&E1	37	138	14	12	201	73 (58, 84)	92 (86, 96)	87 (82, 91)
	A&E2	19	151	18	11	199	51 (34, 68)	93 (88, 97)	85 (80, 90)
	Total	56	289	32	23	400	64 (53, 74)	93 (89, 95)	86 (82, 89)

Table 5.10 *CSR versus radiologist A&E radiograph reporting accuracy*

Body area	% Difference (95% CI)	P-value
A&E1	2 (- 2.2, 7.2)	0.30
A&E2	1 (- 4.0, 6.0)	0.70
Total	1 (- 1.7, 5.2)	0.32

Table 5.11 shows no significant difference in the CSRs' accuracy when reporting A&E1 versus A&E2 radiographs (P=0.97) and can be 95 per cent confident that they report the A&E1 radiographs between 7.0 per cent worse and 7.3 per cent better than A&E2 radiographs. So the accuracy with which the CSRs report the body areas they do not report at present is not statistically different from the body areas they do report. Neither is there significant difference in the radiologists' accuracy when reporting these radiographs (P = 0.63).

Table 5.11 *CSR and radiologist A&E radiograph reporting accuracy*

Profession	Body area	% Difference (95% CI)	P-value
CSR	*A&E1 vs A&E2*	1 (- 7.0, 7.3)	0.97
Radiologists	*A&E1 vs A&E2*	2 (- 5.2, 8.5)	0.63

Some variation in the CSRs' GP radiograph reporting performance is illustrated in Table 5.12, most notably for sensitivity when reporting GP1 radiographs. The 95 per cent CIs for the two CSRs do overlap but are very wide. Table 5.13 shows the CSRs' GP radiograph reporting accuracy is comparable to that of the radiologists and Table 5.14 that it is not significantly different. There is also 95 per cent certainty that radiologists GP radiograph reporting accuracy overall is between 1.1 per cent worse and 8.1 per cent better than the CSRs. Table 5.13 also shows that for all GP radiographs, there is 95 per cent certainty that the CSRs' and radiologists' accuracy fall between 69 per cent to 78 per cent and 72 per cent to 81 per cent respectively, although there is some variation in sensitivity and specificity between the two groups. The CSRs' sensitivity is lower than that of the radiologists when reporting GP1 radiographs but higher for GP2 radiographs. The CIs do overlap but they are wide.

Table 5.12 CSR GP radiograph reporting performance

A	Body area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	GP1	23	43	24	4	94	49 (34, 64)	91 (80, 98)	70 (60, 79)
	GP2	30	47	16	14	107	65 (50, 79)	77 (65, 87)	72 (62, 80)
	Total	53	90	40	18	201	57 (46, 67)	83 (75, 90)	71 (64, 77)
B	**Body area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	GP1	38	33	21	4	96	64 (51, 76)	89 (75, 97)	74 (64, 82)
	GP2	37	42	14	10	103	73 (58, 84)	81 (67, 90)	77 (67, 84)
	Total	75	75	35	14	199	68 (59, 77)	84 (75, 91)	75 (69, 81)

Table 5.13 CSR versus radiologist GP radiograph reporting performance

CSR	Body Area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	GP1	61	76	45	8	190	58 (48, 68)	90 (82, 96)	72 (65, 78)
	GP2	67	89	30	24	210	69 (59, 78)	79 (70, 86)	74 (68, 80)
	Total	128	165	75	32	400	63 (56, 70)	84 (78, 89)	73 (69, 78)
Rad	**Body Area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	GP1	77	68	31	14	190	71 (62, 80)	83 (73, 90)	76 (70, 82)
	GP2	65	97	35	13	210	65 (55, 74)	88 (81, 94)	77 (71, 83)
	Total	142	165	66	27	400	68 (62, 75)	86 (80, 91)	77 (72, 81)

Table 5.14 CSR versus radiologist GP radiograph reporting accuracy

Body area	% Difference (95% CI)	P-value
GP1	4 (-2.3, 10.7)	0.21
GP2	3 (-3.7, 9.4)	0.40
Total	4 (-1.1, 8.1)	0.14

Table 5.15 shows no significant difference in the CSRs' and radiologists' accuracy when reporting GP1 compared with GP2 radiographs, but Table 5.16 shows that any comparison between A&E and GP radiograph reporting accuracy for the two groups yields statistically significant results. The A&E radiographs were consistently reported significantly more accurately than the GP radiographs.

Table 5.15 CSR and radiologist GP radiograph reporting accuracy

Group	Body area	% Difference (95% CI)	P-value
CSR	*GP1 vs GP2*	-2 (-10.9, 6.5)	0.62
Radiologists	*GP1 vs GP2*	-1 (-9.2, 7.5)	0.84

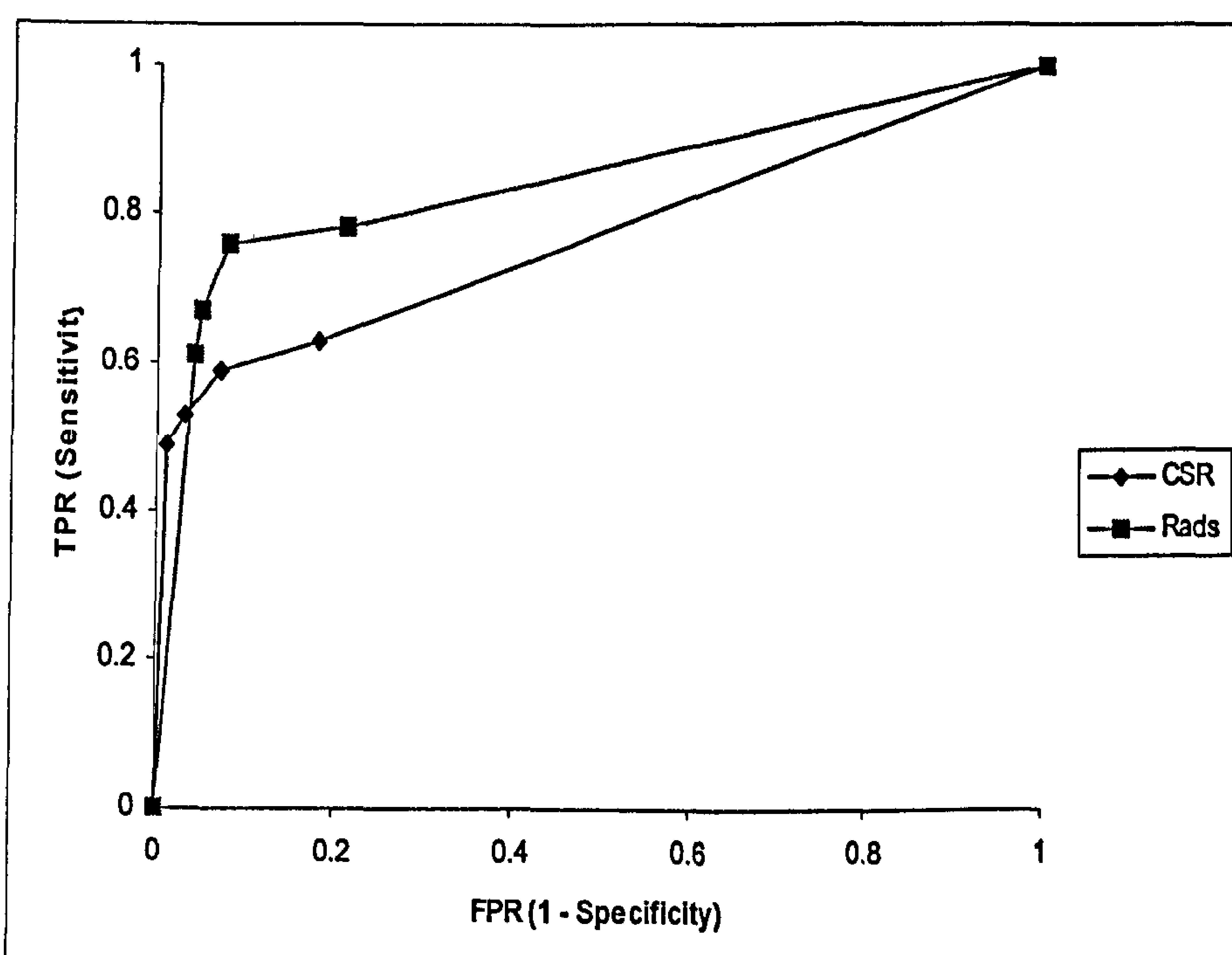
Table 5.16 CSR and radiologist A&E versus GP radiograph reporting accuracy

Group	Body area	% Difference (95% CI)	P-value
CSR	*A&E1 vs GP1*	8 (4.4, 20.6)	0.003
Radiologists	*A&E1 vs GP1*	13 (3.1, 18.4)	0.006
CSR	*A&E1 vs GP2*	11 (2.5, 18.0)	0.010
Radiologists	*A&E1 vs GP2*	10 (2.5, 17.3)	0.009
CSR	*A&E2 vs GP1*	12 (4.2, 20.5)	0.003
Radiologists	*A&E2 vs GP1*	9 (1.3, 17.0)	0.022
CSR	*A&E2 vs GP2*	10 (2.3, 17.9)	0.012
Radiologists	*A&E2 vs GP2*	8 (0.7, 15.8)	0.032
CSR	*A&E total vs GP total*	12 (5.6, 16.9)	<0.001
Radiologists	*A&E total vs GP total*	11 (4.1, 14.9)	<0.001

ROC curves were used to summarise the results of the two groups radiograph reporting performance, exploiting observers' natural tendency for probability scaling when interpreting radiographs and so arguably a more valid reflection of their decision-making (Manning, 1998).

ROC curve plots sensitivity, or true positive rate (TPR), versus 1-specificity, or false-positive rate (FPR), for each category of reports and produces the Az value of the area under the curve. Because the CSRs did not report identical radiographs but a comparable sample, it is not possible to test for a statistically significant difference in their Az values, but this *was* possible for the comparison between CSRs and radiologists.

Figure 5.4 ROC curve comparing CSR and radiologists for A&E1

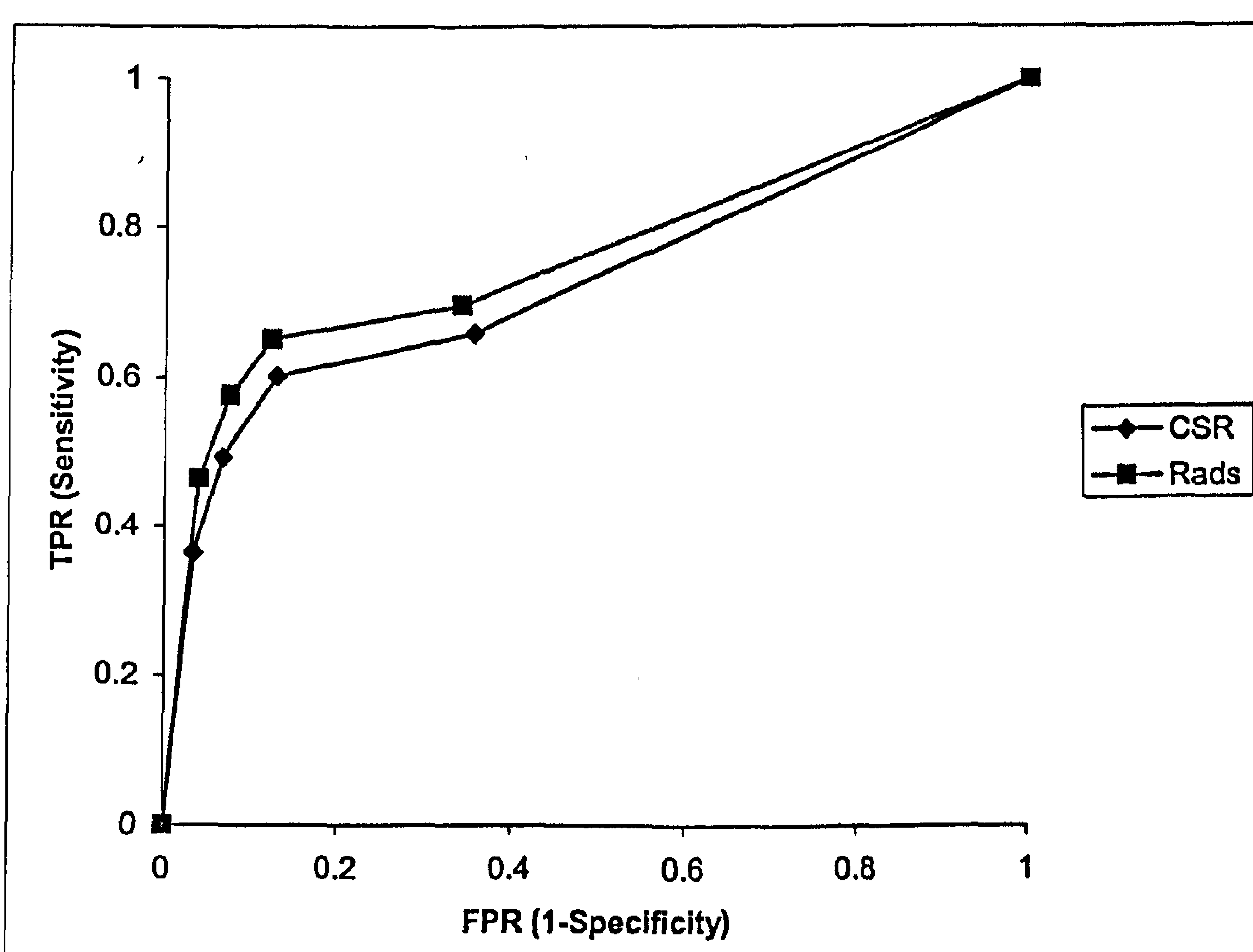


For A&E1 radiographs, the CSRs and radiologists produced mean Az scores of 0.77 and 0.85. Although Figure 5.4 shows the radiologists to be consistently more sensitive, there was no statistically significant difference between their Az values (P=0.09). Table 5.17 shows no significant difference in the performance of each profession when reporting the other three groups. When pooling these groups (n=599), the CSRs and radiologists produced mean Az scores of 0.73 and 0.76 and although Figure 5.5 shows the radiologists to be consistently better for both sensitivity and specificity, this was not significantly different (P=0.17).

Table 5.17 CSR and consultant radiologist Az values

Group	Az value (95% CI)		P-values
	CSR	Radiologists	
A&E1	0.77 (0.69, 0.85)	0.85 (0.78, 0.91)	0.09
A&E2	0.72 (0.62, 0.82)	0.70 (0.60, 0.80)	0.76
GP1	0.70 (0.64, 0.77)	0.74 (0.67, 0.80)	0.37
GP2	0.74 (0.68, 0.81)	0.77 (0.70, 0.83)	0.51
A&E2 + GP1 + GP2	0.73 (0.69, 0.77)	0.76 (0.72, 0.80)	0.17

Figure 5.5 ROC curve comparing CSR and radiologists for all experimental groups



5.8.1.2 Comparison with a single consultant radiologist, but excluding FAR cases

The following three tables present the two professional groups A&E and GP radiograph reporting performance, but exclude cases when the CSR would have liked to seek advice from a radiologist. Tables 5.18 and 5.19 compared with Tables 5.9 and 5.13 show elevated performance in both groups, probably because the excluded cases are likely to be more subtle/difficult. Table 5.18 shows that for group A&E2 the CSRs' accuracy is now 1 per cent higher than the radiologists'. For A&E radiographs overall, the CSRs reported as accurately as the radiologists. Table 5.19 shows that the CSRs' reporting performance for GP radiographs was now even closer to the radiologists'. Table 5.20 shows no statistically significant difference between the CSR and radiologist A&E and GP radiograph reporting accuracy.

Table 5.18 A&E radiograph reporting performance (excl. FAR)

CSR	Body area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	A&E1	29	139	20	8	196	59 (44, 73)	95 (90, 98)	86 (80, 90)
	A&E2	14	144	13	4	175	52 (32, 71)	97 (93, 99)	90 (85, 94)
	Total	43	283	33	12	371	57 (45, 68)	96 (93, 98)	88 (84, 91)
Rad	**Body area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	A&E1	36	135	13	12	196	73 (59, 85)	92 (86, 96)	87 (82, 92)
	A&E2	14	142	13	6	175	52 (32, 71)	96 (91, 98)	89 (84, 93)
	Total	50	277	26	18	371	66 (54, 76)	94 (91, 96)	88 (84, 91)

Table 5.19 GP radiograph reporting performance (excl. FAR)

CSR	Body area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	GP1	60	73	39	4	176	61 (50, 70)	95 (87, 99)	76 (69, 82)
	GP2	43	78	22	13	156	66 (53, 77)	86 (77, 92)	78 (70, 84)
	Total	103	151	61	17	332	63 (55, 70)	90 (84, 94)	77 (72, 81)
Rad	**Body area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	GP1	72	65	28	11	176	72 (62, 81)	86 (76, 93)	78 (71, 84)
	GP2	43	81	23	9	156	65 (52, 76)	90 (82, 95)	79 (72, 86)
	Total	115	146	51	20	332	69 (62, 76)	88 (82, 92)	79 (74, 83)

Table 5.20 CSR versus radiologist radiograph reporting accuracy (excl. FAR)

Body area	% Difference (95% CI)	P-value
A&E1	1 (-3.0, 6.1)	0.52
A&E2	-1 (-5.6, 3.3)	0.62
A&E Total	0 (-2.9, 3.5)	0.88
GP1	2 (-4.4, 8.9)	0.50
GP2	1 (-4.8, 8.7)	0.61
GP Total	2 (-2.6, 6.9)	0.38

5.8.1.3 Comparison with a double blind consultant radiologist

The following three tables present the two professional groups A&E and GP radiograph reporting performance, compared with a double blind consultant radiologist report, i.e. when the reference standard report agreed with the report of a consultant radiologist from the study presented in Chapter 4. This analysis includes cases when the CSRs would have liked to seek advice from a radiologist.

Tables 5.21 and 5.22 show an increase in the indices of performance for the two professional groups when compared with this new reference standard, suggesting that the two independent radiologist reports used to generate this reference standard were concordant for unequivocal and therefore less difficult cases. There are otherwise similar percentage differences in accuracy between the two groups. Table 5.23 shows no significant difference between the CSR and radiologist A&E and GP radiograph reporting accuracy.

Table 5.21 A&E radiograph reporting performance (double blind)

CSR	Body area	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	A&E1	20	75	6	7	108	77 (56, 91)	91 (83, 96)	88 (80, 93)
	A&E2	11	135	5	11	162	69 (41, 89)	92 (87, 96)	90 (84, 94)
	Total	31	210	11	18	270	74 (58, 86)	92 (88, 95)	89 (85, 93)
Rad	**Body area**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	A&E1	21	76	5	6	108	81 (61, 93)	93 (85, 97)	90 (83, 95)
	A&E2	11	136	5	10	162	69 (41, 89)	93 (88, 97)	91 (85, 95)
	Total	32	212	10	16	270	76 (61, 88)	93 (89, 96)	90 (86, 94)

Table 5.22 GP radiograph reporting performance (double blind)

CSR	Body areas	TP	TN	FN	FP	Total	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
	GP1	41	57	14	5	117	75 (61, 85)	92 (82, 97)	84 (76, 90)
	GP2	45	79	10	20	154	82 (69, 91)	80 (71, 87)	81 (73, 86)
	Total	86	136	24	25	271	78 (69, 85)	84 (78, 90)	82 (77, 86)
Rad	**Body areas**	**TP**	**TN**	**FN**	**FP**	**Total**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
	GP1	49	53	7	8	117	88 (76, 95)	87 (76, 94)	87 (80, 93)
	GP2	42	87	15	10	154	74 (60, 84)	90 (82, 95)	84 (77, 89)
	Total	91	140	22	18	271	81 (72, 87)	89 (83, 93)	85 (80, 89)

Table 5.23 CSR versus radiologist radiograph reporting accuracy (double blind)

Body area	% Difference (95% CI)	P-value
A&E1	2 (-4.4, 8.1)	0.56
A&E2	1 (-4.4, 5.6)	0.81
A&E Total	1 (-2.8, 5.0)	0.58
GP1	3 (-4.8, 11.6)	0.41
GP2	2 (-4.0, 10.5)	0.38
GP Total	3 (-2.1, 8.8)	0.23

5.8.2 *Diagnostic Outcome*

This following presents the results of the effect of the CSR and radiologist incorrect reports on the referring clinician's diagnosis and confidence in that diagnosis. To calculate the percentage change in diagnosis, the number of diagnoses the clinician changed based on incorrect reports was divided by the total number of diagnoses based on incorrect reports and then multiplied by 100. To calculate the effect on confidence in diagnosis, the pre-report confidence figures were deducted from the confidence figures when the report was available. The percentage difference between these figures was summated and divided by the sample size to produce the mean percentage change in confidence. The first section presents the results about changes in the clinician's diagnosis and the second section changes in their confidence in that diagnosis.

5.8.2.1 Changes in the diagnosis of the clinician

To assess changes in diagnosis, a consultant radiologist judged whether the pre-report diagnosis was the same as when the report was available. The radiologist did this with access to clinical details but blind to who made the report.

Table 5.24 shows that for A&E1 radiographs the clinicians' diagnosis changed in 39 per cent and 50 per cent of cases respectively after reading incorrect CSR and radiologist reports, though this 11 per cent difference in change of diagnosis has quite low precision. There is 95 per cent certainty that the clinicians changed their diagnosis based on CSR report for A&E1 radiographs between 35.8 per cent less and 14.5 per cent more than for a radiologist report. The opposite occurred with A&E2 radiographs, so that for all A&E radiographs there was only a 1 per cent difference in cases when clinicians changed their diagnoses based on incorrect CSR or radiologist reports.

Table 5.24 Changes in diagnoses for A&E patients

Body area	Profession	Change (%)	% Difference (95% CI)
A&E1	CSR	12/31 (39)	-11 (-35.8, 14.5)
	Rads	13/26 (50)	
A&E2	CSR	11/31 (36)	8 (-15.9, 30.7)
	Rads	8/29 (28)	
Total	CSR	23/62 (37)	-1 (-18.6, 16.3)
	Rads	21/55 (38)	

Table 5.25 shows that for GP1, GP2 and combined, the GPs' diagnoses changed 9 per cent, 15 per cent and 12 per cent more often using radiologist rather than CSR incorrect reports. The CIs for the percentage difference cross zero, suggesting no significant difference, but the width of the CI suggests a lack of precision in the estimate of the true population value. In particular, for GPs in total, the upper limit of the CI only just crosses zero.

Table 5.25 Changes in diagnosis for GP patients

Body area	Profession	Change (%)	% Difference (95% CI)
GP1	CSR	22/53 (42)	-9 (-28.7, 10.2)
	Rads	23/45 (51)	
GP2	CSR	11/54 (20)	-15 (-32.2, 2.4)
	Rads	17/48 (35)	
Total	CSR	33/107(31)	-12 (-25.3, 1.2)
	Rads	40/93 (43)	

5.8.2.2 Changes in the diagnostic confidence of the clinician

The next two tables present findings on changes in clinician's confidence in diagnosis after seeing the incorrect report of a CSR or radiologist.

Table 5.26 shows, for example, that the referring clinicians' confidence in their diagnoses increased by 31 per cent and 44 per cent based on the incorrect A&E1 report of a CSR and radiologist respectively. This reflects a 13 per cent mean difference in the confidence of the clinicians. There is also 95 per cent certainty that the mean difference in the clinicians' confidence in their diagnoses based on an incorrect CSR report is between 29.0 per cent less and 3.6 per cent more than for an incorrect radiologist report. Although the CI do cross zero, there is a suggestion that the clinicians' confidence increased more based on an incorrect radiologist report.

Table 5.26 Changes in diagnostic confidence for A&E patients

Body area	Profession	% Change (95% CI)	% Difference (95% CI)
A&E1	CSR	31 (20, 43)	-13 (-29.0, 3.6)
	Rads	44 (33, 55)	
A&E2	CSR	18 (6, 31)	-5 (-21.1, 11.0)
	Rads	23 (13, 34)	
Total	CSR	25 (16, 33)	-8 (-19.8, 3.5)
	Rads	33 (25, 41)	

Table 5.27 shows similar findings for changes in clinicians' confidence based on CSR and radiologist reports of GP plain radiographs, with a smaller percentage change based on CSR incorrect reports. In particular, for GP radiographs in total, the lower CI does not cross zero, suggesting that the clinicians' confidence in their diagnoses of GP patients changed significantly more based on incorrect reports of the radiologists.

It is also interesting to note that clinicians' increase in diagnostic confidence is considerably lower when based on reports of GP radiographs. Indeed, having access to the incorrect CSR reports produced almost no percentage change in the clinicians' confidence in their diagnoses.

Table 5.27 Changes in diagnostic confidence for GP patients

Body area	Profession	% Change (95% CI)	% Difference (95% CI)
GP1	CSR	0 (-8, 9)	-6 (-18.2, 6.7)
	Rads	6 (-3, 16)	
GP2	CSR	6 (-2, 13)	-10 (-21.5, 0.3)
	Rads	16 (8, 24)	
Total	CSR	3 (-2, 8)	-8 (-16.5, -0.1)
	Rads	11 (5, 18)	

5.8.3 Therapeutic Outcome

This section presents the results of the effect of CSR and radiologist incorrect reports on the referring clinicians' confidence in choice of patient management. To calculate the effect on therapeutic confidence, the pre-report confidence figures were deducted from the confidence figures when the report was available. The percentage difference between these figures was summated and divided by the sample size to produce the mean percentage change in confidence.

Tables 5.28 and 5.29 present changes in the confidence clinicians had in their management after seeing the incorrect report of a CSR or radiologist. Table 5.28 shows some variation in the mean percentage difference in the clinicians' confidence based on the two professions incorrect reports for A&E1 and A&E2 patients. However, for A&E in total, the clinicians' confidence in their management increased by 29 per cent and 27 per cent for incorrect reports of CSR and radiologists respectively - a mean difference of 2 per cent. There was 95 per cent certainty that the clinicians' confidence in their management based on an incorrect CSR report was between 10.0 per cent less and 13.6 per cent more than for an incorrect radiologist report. The small percentage difference and CI obviously crossing zero, suggest that there is no clear evidence of a significant difference.

Table 5.28 Changes in therapeutic confidence for A&E patients

Body area	Profession	% Change (95% CI)	% Difference (95% CI)
A&E1	CSR	31 (23, 43)	-7 (-22.5, 8.1)
	Rads	38 (27, 50)	
A&E2	CSR	25 (11, 39)	8 (-9.8, 25.6)
	Rads	17 (6, 28)	
Total	CSR	29 (20, 37)	2 (-10.0, 13.6)
	Rads	27 (19, 35)	

Findings indicate that the clinician's confidence in their management of GP patients changed more based on incorrect radiologist reports (Table 5.29). Notably, for GP radiographs in total, the clinicians' confidence in their management increased by 12 per cent and 20 per cent based on the incorrect report of CSRs and radiologists respectively - a mean difference of 8 per cent. There was 95 per cent certainty that the clinicians' confidence in their management based on an incorrect CSR report was between 14.0 per cent and 1.4 per cent less than for an incorrect radiologist report. Because the CI does not cross zero, this indicates a significant difference in the clinicians' confidence in their choice of management based on incorrect CSR and radiologist reports of GP plain radiographs.

Table 5.29 Changes in therapeutic confidence for GP patients

Body area	Profession	% Change (95% CI)	% Difference (95% CI)
GP1	CSR	8 (3, 13)	-7 (-14.1, 0.3)
	Rads	15 (10, 20)	
GP2	CSR	16 (10, 22)	-8 (-18.7, 2.1)
	Rads	24 (15, 33)	
Total	CSR	12 (8, 16)	-8 (-14.0, -1.4)
	Rads	20 (15, 25)	

The findings from Tables 5.28 and 5.29 also show a lower percentage increase in clinicians' confidence based on incorrect reports of GP compared with A&E plain radiographs. As an example, for A&E radiographs in total, the mean increase in confidence based on incorrect CSR reports was 29 per cent compared with 12 per cent for GP radiographs in total. The upper limit of the CI for GP radiographs (i.e. 16 per cent) does not cross the lower limit of the CI for A&E radiographs (i.e. 20 per cent), suggesting a significant difference in the mean percentage change in confidence based on incorrect CSR reports of A&E and GP radiographs.

5.8.4 Patient Outcome

This final section presents findings on the effect of CSR or radiologist incorrect reports compared with the reference standard resulting in clinically important differences in management that could affect patient outcome. The different outcomes in the following tables were defined in Table 5.7.

The most important finding in Table 5.30 is that when a CSR and radiologist report A&E1 radiographs incorrectly, in 35 per cent and 38 per cent of cases respectively this may have resulted in a clinically important difference in patient management. However, this is only a 3 per cent difference in cases between the two professional groups.

Table 5.31 shows that for A&E2 this difference is only 1 per cent. The same applies for Table 5.32 and the findings for A&E in total. Despite the lack of precision, these findings do indicate that there is no clear evidence of a significant difference in the effect of incorrect CSR and radiologist reports of A&E plain radiographs on patient outcome.

Table 5.30 Patient outcome for A&E1

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	11/31 (35)	8 (-16.1, 31.6)
	Rads	7/26 (27)	
Clinically unimportant	CSR	9/31 (29)	-6 (-29.7, 18.4)
	Rads	9/26 (35)	
Clinically important	CSR	11/31 (35)	-3 (-27.8, 21.7)
	Rads	10/26 (38)	

Table 5.31 Patient outcome for A&E2

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	19/31 (61)	2 (-21.8, 26.9)
	Rads	17/29 (59)	
Clinically unimportant	CSR	1/31 (3)	-4 (-19.4, 10.4)
	Rads	2/29 (7)	
Clinically important	CSR	11/31 (35)	1 (-23.0, 24.7)
	Rads	10/29 (34)	

Table 5.32 Patient outcome for A&E in total

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	30/62 (48)	4 (-13.3, 22.4)
	Rads	24/55 (44)	
Clinically unimportant	CSR	10/62 (16)	-4 (-18.5, 10.2)
	Rads	11/55 (20)	
Clinically important	CSR	22/62 (35)	-1 (-18.3, 16.3)
	Rads	20/55 (36)	

The most important finding in Table 5.33 is that in 38 per cent and 51 per cent of cases respectively, when a CSR and radiologist report GP1 radiographs incorrectly this may have a clinically important affect on patient management. The difference of 13 per cent of cases does not appear to be significant, as the CIs cross zero, although it is wide reflecting low power to detect a difference.

Table 5.33 Patient outcome for GP1

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	22/53 (42)	26 (8.1, 42.2)
	Rads	7/45 (16)	
Clinically unimportant	CSR	11/53 (21)	-12 (-30.2, 5.0)
	Rads	15/45 (33)	
Clinically important	CSR	20/53 (38)	-13 (-32.2, 6.4)
	Rads	23/45 (51)	

The same applies to Tables 5.34 and 5.35, with a fewer percentage of cases based on CSR incorrect reports for GP2 and GP radiographs in total resulting in clinically important differences in patient management. Again, the CIs around the difference in percentage of cases cross zero, suggesting that there is no significant difference between the professional groups.

Table 5.34 Patient Outcome for GP2

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	23/54 (43)	18 (-1.0, 34.9)
	Rads	12/48 (25)	
Clinically unimportant	CSR	11/54 (20)	-9 (-25.7, 8.0)
	Rads	14/48 (29)	
Clinically important	CSR	20/54 (37)	-9 (-27.4, 10.3)
	Rads	22/48 (46)	

Table 5.35 Patient outcome for GP in total

Outcome	Profession	Change (%)	% Difference (95% CI)
No difference	CSR	45/107 (42)	22 (8.8, 33.7)
	Rads	19/93 (20)	
Clinically unimportant	CSR	22/107 (21)	-10 (-22.8, 1.5)
	Rads	29/93 (31)	
Clinically important	CSR	40/107 (37)	-11 (-24.4, 2.7)
	Rads	45/93 (48)	

5.9 Discussion

Findings presented in this chapter show that there is no clear evidence to claim a statistically significant difference in the accuracy of CSRs and consultant radiologists when reporting A&E and GP plain radiographs. Furthermore, the CIs around the CSR and radiologist estimates of sensitivity and specificity for different body areas and patient types overlap each other, though the results of the ROC analyses were not so conclusive. Notably, when the two professional groups reported A&E1 radiographs (those that the CSRs currently report), the P-value approached statistical significance (P=0.09). The ROC curve in Figure 5.6 also showed that for experimental groups combined, the radiologists were consistently better for both sensitivity and specificity, but again with no clear evidence to suggest a significant difference (P=0.17). It was also interesting to find that both professional groups' sensitivity was lower than their specificity when interpreting A&E and GP referred radiographs, in accord with what was found in Chapter 4, and again probably explained by the criteria used to judge reports for concordance. Similarly, the reporting performance of GP radiographs was also significantly lower than for A&E radiographs for both professional groups.

In the case of incorrect CSR or radiologist reports inappropriately changing a clinicians' diagnoses for A&E and GP patients, no clear evidence emerged to suggest a significant difference between the two groups. The high percentage change in clinicians' diagnoses based on incorrect reports is interesting. Clinicians changed their diagnoses for A&E1 and GP1 radiographs for around 50 per cent and 40 per cent of cases based on incorrect radiologist and CSR reports respectively. Reports of these radiographs would appear to be important to clinician decision-making justifying the need to assess their effect on clinician diagnosis.

No clear evidence was found to suggest a significant difference in the confidence of the A&E Specialist Registrar's diagnosis based on CSR and radiologist incorrect reports. However, for GP radiographs in total, there was evidence that clinicians' confidence in their diagnoses changed significantly based on incorrect radiologist reports. The lower percentage changes in confidence based on incorrect reports of GP radiographs compared with A&E radiographs is not surprising, considering the CSR and radiologist lower reporting performance for these radiographs. It might be that because GP radiographs are more difficult to report, it influences the confidence with which the radiograph appearances are reported and subsequently the confidence of the referring clinician. Furthermore, the finding that clinicians' confidence in their diagnoses diminished based on CSR incorrect reports suggests a difference in how CSRs convey information - as also reflected in the ROC analyses. Neither is this surprising, as the CSRs do not currently report these radiographs in normal practice, so are probably more uncertain and cautious about what to report. Finally, these findings question why the GP would refer the patient for a radiograph if the reports are not changing the confidence in their diagnoses. It would have been interesting to see the changes in confidence as a result of correct reports.

Evidence was not found to support the significance of the mean percentage difference in the clinician's choice of management, based on the incorrect reports of either professional group for A&E patients. However, findings did suggest a significant difference for GP patients in total, with the clinicians' confidence in management being lower when based on the CSRs' incorrect reports. These findings are comparable with changes in confidence in their diagnoses. Overall, the findings suggest that changes in the clinicians' confidence in their management plans showed a greater increase based on incorrect CSR and radiologist reports of A&E rather than GP radiographs.

Perhaps most importantly, there was no evidence to suggest a significant difference in cases, based on CSRs' and radiologists' incorrect reports of A&E and GP radiographs compared with the reference standard, that would result in clinically important differences in patient management. In particular, although the difference in GPs' confidence in diagnosis and management plans was significantly higher based on radiologist incorrect reports, this did not translate into a higher percentage of cases that may have adversely affected patient outcome.

Certain limitations to the study should be considered. First, although the design meant that the findings are attributable to the two professional groups performance, it does not entirely reflect clinical practice. For example, the two professional groups could not discuss radiographs with colleagues or access previous reports as would normally occur. Nor could it be guaranteed that colleagues would not discuss radiographs with each other, or search for the previous reports. Because the CSRs and radiologists reported the radiographs in the knowledge that they were under scrutiny, their reporting behaviour might have been changed. Moreover, as the study was not conducted during clinical practice, they knew their reports would not affect patient management and outcome, leaving the study susceptible to the Hawthorne effect, which occurs when the knowledge of being under evaluation influences behaviour (Last, 1995). However, Table 4.3 in Chapter 4 shows that A&E radiograph reporting accuracy before and after the introduction of the CSRs was 87 per cent and 86 per cent respectively. Table 5.9 in Chapter 5 illustrates that CSRs' and radiologists' reporting accuracy for the same sample of A&E radiographs was 85 per cent and 86 per cent respectively. Table 4.5 in Chapter 4 shows that radiologist reporting accuracy for GP radiographs was 76 per cent both before and after the introduction of the CSR. Table 5.13 in Chapter 5 shows that CSR and radiologist reporting accuracy for the same sample of GP radiographs was respectively 73 per cent and 77 per cent. So reporting accuracy of the two professional groups was almost identical when interpreting the same sample of radiographs during clinical practice and under controlled conditions, suggesting that the CSRs and radiologists did not change their behaviour when reporting radiographs for this study.

Selection bias could be a potential criticism. Block randomisation, stratified by pertinent body areas, was not employed for the selection of radiographs interpreted by the consultant radiologists during audit. This is because it was not logistically possible to ensure which radiologists would be available and how much time they could devote to this activity. So the author, who organised these sessions, could only select a consecutive series of between 20 to 30 radiographs for whichever radiologist was available. No explicit bias in radiograph selection was present, as neither the radiologist nor the author selectively chose the radiographs. Each radiologist reported a convenience sample of the stratified random sample.

Assessment of the effects of incorrect reports reflected merely what *might* have happened. Patient history provided to clinicians was only from the original X-ray request form. Furthermore, the assessment of how incorrect CSR or radiologist reports, compared with the reference standard, could affect patient management and outcome was based on the subjective judgement of the clinician. Again, this was the only feasible method, as the study had to be designed around the reporting of a retrospective sample of radiographs. The sample sizes during the various stages of assessing performance were also small, so the study has low power to detect significant differences.

Finally, the study has limited generalisability. It was based at a single District General Hospital in a rural area, and involved only a few health care professionals, in particular only two CSRs. The results apply to the level of performance of selectively trained radiographers in this setting.

Nevertheless, this study is designed more rigorously than the studies included in the systematic review presented in Chapter 2, with various precautions for prevention of bias, so that the reporting performance and subsequent effects could be attributed to the two professional groups. This included independence in the reporting of radiographs by all professionals and the reference standard, application of the reference standard to all radiographs, and blinding the arbiter and clinicians to information that could adversely affect their decision-making. The study was also specifically designed to assess the efficacy of CSRs and radiologists reporting plain radiographs under controlled conditions, to avoid the reports having a direct effect on patient management and outcome. The study was designed to assess the hypothetical effect of CSR and radiologist reports before deciding on the potential or future need for a larger and appropriately powered study to assess their performance during clinical practice. Evidence is provided here about CSR and radiologist reporting both A&E and GP plain radiographs and unlike previous studies, presents the subsequent effects on clinician decision-making and its translation into effect on patient outcome.

5.10 Conclusion

This study found no significant difference between the CSR and consultant radiologist performance when reporting plain radiographs for A&E and GP patients. Some findings *did* suggest that the subsequent effect of incorrect reports by radiologists adversely influenced GPs' confidence in their diagnosis and management plans significantly more than the CSRs' incorrect reports. But ultimately there was no evidence of a significant difference in the percentage of cases that an incorrect CSR and radiologist report might affect patient outcome. The need for an appropriately powered multi-centre study to assess the cost-effectiveness of selectively trained radiographers reporting GP plain radiographs has justification in the findings of this study.

Having presented the results of the two primary studies about the effect of CSR and radiologist reporting plain radiographs using both pragmatic and explanatory designs, it was now appropriate to present the analysis of the *cost* of CSR reporting A&E plain radiographs, as discussed in the next chapter.

Chapter 6

A cost analysis of A&E plain radiograph reporting

6.1 Introduction

The introduction of the NHS and Community Care Act in June 1990 resulted in a major re-structuring of the funding and delivery of health care, with particular emphasis placed on resource management and the need for improved efficiency. This promoted the more flexible and creative use of allied health care professionals, as skill mix could potentially make substantial savings on labour costs (Kletzenbauer, 1996). A recent review identified many issues which remain unresolved regarding the substitution of non-medical professionals for doctors (Richardson et al, 1998). Absence of appropriate research and the reluctance of managers and other decision makers to practice evidence based policy are contributory causes. Richardson made an urgent recommendation for further research to evaluate the cost-effectiveness of these initiatives.

As discussed in Chapter 1, the creative use of selectively trained radiographers as a substitute for radiologists in reporting plain radiographs, was seen as a skill mix initiative capable of improving the efficiency of an X-ray department. The crude assumption is that employing a radiographer to report is less costly than employing a radiologist. Indeed, the College of Radiographers (CoR) heralded this skill mix as having the capacity to revolutionise the cost-effective management of the patient in clinical radiology and other imaging dependent services (CoR, 1997). The reality is more complex. For example, are the outputs (such as ensuring that all radiographs continue to be reported) the same for both selectively trained radiographers and radiologists?

Whether the efficiency of a reporting service can be improved by selectively trained radiographer reporting also depends on what they would otherwise have been doing with their time while training and subsequently reporting radiographs. This is the 'opportunity cost' of radiographer reporting. Management of resources available as a consequence, for example, of radiographer reporting freeing radiologists to make better use of their time and expertise, is another important factor. People, time, facilities, equipment and knowledge are scarce, so radiographers' and radiologists' time and skills must be managed productively.

Very limited evidence is available on the potential for cost savings from radiographer reporting, as found in the systematic review presented in Chapter 2. Just one study was identified that explored the cost of implementing radiographer reporting accident and emergency (A&E) radiographs. Piper et al (2000) found that the additional costs for providing a radiographer reporting service in four NHS trusts ranged from nil to £15000 per annum, with radiographers spending around 0.5 whole time equivalent (wte) reporting. However, only a minimal amount of cost data was collected, addressing staff costs. The report acknowledged that the cost data was not reliable and that further evaluation was required.

Chapter 4 presented results in terms of accuracy and subsequent effects on patient management and outcome, of the introduction of Clinical Specialist Radiographers (CSRs) reporting A&E plain radiographs of the appendicular skeleton at Trust A - the first primary study. In the absence of robust evidence about the cost of selectively trained radiographers reporting plain radiographs, it was timely to address whether the introduction of the CSRs to the reporting service at Trust A was less costly than the radiologists' reporting. The *primary* objective of this study was to examine the cost of introducing CSR reporting A&E radiographs of the appendicular skeleton at Trust A.

Chapter 5 presented the results of the second primary study, examining CSRs and consultant radiologists reporting A&E and general practitioner (GP) plain radiographs at Trust A. The cost of extending their role to include the remaining body areas for A&E patients was also explored as a *secondary* objective to this study, but too many assumptions would have to be made to analyse the cost of CSRs also reporting general practitioner (GP) plain radiographs, so it was not included as an objective. It was also more appropriate to focus on the cost of radiographer reporting A&E plain radiographs, reflecting the current change in clinical practice.

To supplement the discussion about whether the introduction of CSR reporting A&E radiographs at Trust A improved service efficiency, possible resource management implications were addressed, using the results of a survey at Trust A on the acceptability of CSR reporting to different professionals (Brealey et al, 2002).

The study presented in this chapter should help managers decide on the benefit of investing in selectively trained radiographers to report A&E plain radiographs.

6.2 Objectives

6.2.1 Primary objective

- To examine the cost of CSR reporting A&E radiographs of the appendicular skeleton compared with radiologist reporting.

6.2.2 Secondary objective

- To explore the cost of CSR reporting A&E radiographs for all body areas compared with radiologist reporting.

6.3 Methods

The method of economic evaluation used to examine the cost of CSRs and radiologists reporting A&E plain radiographs is presented here, followed by a discussion about the costs considered suitable for inclusion in the analysis of CSR and radiologist reporting and how they were measured and valued.

6.3.1 Choice of economic evaluation

Health care evaluation has been described as 'the process of choosing between alternative health care policies by estimating the net value of each' (Russell, 1983). Such an estimate is achieved by identifying the inputs consumed and outputs generated. To explore the relationship between these costs and consequences requires an economic evaluation, of which there are various types (Drummond et al, 1987).

A 'cost analysis' would focus solely on the cost of CSR and radiologist reporting - not the consequences, representing only a partial form of economic appraisal. However when the consequences of two or more alternatives, such as CSR or radiologist reporting, are shown to be equivalent, the study can be termed a 'cost-minimisation' study. A 'cost-effectiveness analysis' measures the consequences of CSR or radiologist reporting in the most appropriate natural or physical units, such as 'cases correctly diagnosed'. No attempt is made to value the consequences so implicitly it is assumed that the output concerned is in some sense worth having. In 'cost-utility analysis' the consequences of CSR or radiologist reporting would incorporate the notion of value and be measured in time units adjusted by health utility weights. The effect of the report could be presented not just as the number of years a patient survives, which is implicitly something worth having, but would value those years in terms of patient quality of life. Patients would be asked to complete an instrument like the EuroQol (EuroQol Group, 1991) to measure differences in utility depending on who reported the radiograph. 'Cost-benefit analysis' would attempt to value the consequences of CSR or radiologist reporting in financial terms to make them commensurate with the costs, requiring the use of techniques like 'willingness-to-pay'. Patients would be asked to state what they would be willing to pay for their radiographs to be reported by CSRs rather than radiologists, bearing in mind the possible difference in consequences depending on who reports the radiographs.

It was apparent from the outset that the latter two methods of economic evaluation would not be feasible. Both would require patient-completed questionnaires, but the studies presented in Chapters 4 and 5 collected data about CSR and radiologist reporting using a retrospective sample of radiographs and reports. The prospective collection of data from a similar sample of patients was not possible in the context of the other objectives of the thesis, so the economic evaluation would be a cost-minimisation or cost-effectiveness analysis.

Chapters 4 and 5 presented the two primary studies that included an assessment of CSRs and radiologists reporting A&E plain radiographs at Trust A. Evidence from Chapter 4 showed no statistically or clinically significant difference in A&E radiograph reporting performance following the introduction of the CSRs. Neither was there evidence to suggest a statistically significant difference in A&E patient management, nor any obvious difference in patient outcome in terms of patient re-attendance to A&E or the X-ray department. Study results from Chapter 5 provided no clear evidence to support a claim that consultant radiologists reported A&E plain radiographs significantly more accurately than the CSRs, nor that there was a significant difference in how CSR and radiologist incorrect reports of A&E plain radiographs would affect the referring clinician's choice of diagnosis and confidence in that diagnosis, confidence in their management plans, and patient outcome. Additionally, a conclusion from the systematic review in Chapter 2 found no clear evidence to claim that selectively trained radiographers report A&E plain radiographs less accurately than radiologists of varying seniority.

Evidence from these preceding chapters suggested that CSRs' and radiologists' A&E radiograph reporting performances were similar enough to assume that there would be no variation in consequences to the NHS or patients. In the absence of evidence of a wider impact, this would indicate that a narrow perspective of the NHS could be taken - the X-ray department at Trust A. Discussion is now confined to the cost to the X-ray department of CSR and radiologist reporting, as the consequences were assumed to be the same.

6.3.2 *Identification of costs of plain radiograph reporting*

A variety of costs are considered in an economic evaluation, including variable costs (e.g. time of health professionals), fixed or overhead costs (e.g. light, heat, rent, salaries or capital costs), and out-of-pocket expenses incurred by patients and/or family members (e.g. time lost from work). Other costs include the anxiety or pain associated with a treatment experienced by patients. The first task was to identify the costs that should be included in the analysis of CSR and radiologist reporting. The costs excluded from the analysis are then presented.

6.3.2.1 Costs included in the analysis of CSR and radiologist reporting

The costs associated with CSR reporting included capital costs such as purchasing and installing new equipment for CSRs to use when reporting (e.g. computer, printer, viewing boxes, chair, worktops). Training costs included registration on a MSc course and associated travel expenses. Overhead costs (e.g. heat, light) were those associated with the rooms used by CSRs, radiologists and secretaries for reporting activity. Time spent by CSR or radiologists reporting radiographs and the CSRs or secretaries typing up reports must also be considered. There were no additional medico-legal or insurance costs to take into account.

6.3.2.2 Costs excluded from the analysis of CSR and radiologist reporting

While the two CSRs attended an MSc course in image interpretation at Leeds, they were employed as 1.0 wte but worked only 0.5 wte. On completion of their training, they spent 0.5 wte reporting and the other 0.5 wte performing normal duties (e.g. X-ray, fluoroscopy, general administration). A 1.0 wte basic radiographer was employed to provide cover for the CSRs during their training and subsequent reporting in clinical practice. However, the employment of this additional radiographer can not be directly related to the cost of the intervention (CSR reporting) without a study to examine the work of all activities in the department before and after the CSRs were introduced to the reporting service. Instead, a more direct observational study was conducted to identify the resources required for the two alternatives: CSR and radiologist reporting the same workload (e.g. radiographs). Resources required for the CSR to perform the reporting activity are: the cost of training the CSR to report; provision of a room and equipment; time taken for reporting. To employ an additional radiographer was how the X-ray department decided to use resources so that the normal duties no longer performed by the CSR were still undertaken.

Cost of consumables, such as bulbs for viewing boxes, paper for reports and print ribbons, were also excluded. These are negligible costs and are the same for both professions. Equipment used by secretaries to type reports dictated to tape by radiologists is an additional cost of radiologist reporting. This cost was excluded for two reasons: the cost of the dictation equipment (a few hundred pounds) was small; it was used not just for typing radiologist reports of A&E plain radiographs but for all other types of patient, and for different imaging modalities (e.g. Ultrasound, Computed Tomography, Magnetic Resonance Imaging). In the context in which it was used the cost became negligible.

The potential costs of auditing CSR reporting during normal clinical practice to assure quality were also excluded. Protected time ensures that during normal clinical practice audits of any aspect of the service provided by an X-ray department can be performed. X-ray departments often monitor the accuracy of plain radiograph reporting by radiologists, so assuring CSRs reporting during audit was to include them in a normal activity. The opinion of the radiologist from whom the CSR seek advice about A&E radiographs was that they ask for advice once every two weeks, which equates to around 25 plain radiographs a year. In 2000/1 the radiographers reported close to 10,000 radiographs, so seeking advice for 25 radiographs is equivalent to 0.25 per cent of all radiographs and a negligible additional cost. Patient out-of-pocket expenses, or time lost from work, were not relevant as radiographs are not reported when the patients are at the hospital. Variation in the time CSRs or radiologists report would not affect, for example, the length of time the patient and possibly a family member have to wait in the hospital for the report.

6.3.3 *Measurement of costs of plain radiograph reporting*

Having identified the costs of CSR or radiologist reporting A&E plain radiographs to the X-ray department, it was necessary to decide on the units of measurement.

Costs of CSR reporting include the capital costs, which can be measured in terms of numbers of computers, printers, viewing boxes, chairs, or worktops purchased. Training costs are measurable in terms of the number of CSRs registered on the MSc course and the number of times travel was required to attend the course. Staff costs of the CSRs can be measured in terms of the seconds it takes to report a radiograph and type the report. Radiologist reporting costs would be measured in similar units. Overhead costs were measured in number of square metres.

6.3.4 *Valuation of costs of plain radiograph reporting*

Costs identified were based on exact costs and best available estimates provided by the X-ray department at Trust A. All were valued in units of local currency, i.e. pounds sterling, for the base year 2001/2. Gross Domestic Product (GDP) inflation indices were used for the capital and training costs and the Hospital & Community Health Services (HCHS) for staff costs (Netten & Curtis, 2002 p.187).

The capital costs associated with CSR reporting (e.g. computer, printer, changes to the building) represent an investment in CSR reporting at a single point in time. They also represent an investment used over time, which will wear out and depreciate. Capital cost has two components: the opportunity cost of the funds tied up in the capital asset and representing a lost opportunity for investment in another service; depreciation over time of the asset itself. The method used to measure and value the capital costs was to annuitize the initial capital outlay over the useful life of the asset, that is to calculate the equivalent annual cost (Drummond et al, 1997). Both the depreciation and opportunity cost of the capital cost is automatically incorporated in this method.

Salary of the staff involved in reporting represented a further cost. Clinical time spent reporting can be observed, but the two professional groups may have a different potential to be clinically productive if they work a different number of hours a week, vary the time dedicated to clinical work or administration, and perform different activities. It is essential to adjust for this when valuing the time a CSR or radiologist spends reporting radiographs. This in turn will be reflected in the cost of the two professional groups reporting. The following describes how time was valued in 'clinical minutes'.

An audit of radiologists' workload at Trust A, excluding on call, found an average working week to be 38 hours. Four of these hours, i.e. about 10 per cent, were allocated to administration, such as clinical audit, making the radiologists' clinical activity 34 hours, or 90 per cent, of their working week. The CSR had an average working week of 35 hours. Fifty per cent of their time was allocated to reporting and of the other time around 50 per cent was spent on clinical work (e.g. X-ray, fluoroscopy) and 50 per cent on administration. CSR clinical activity was 26 hours per week, or 75 per cent of their work. In the knowledge that radiologists spend more time doing clinical work than CSRs, it was possible to weight the clinical minute of a radiologist relative to the clinical minute of a CSR. As the radiologists spend 34 hours a week on clinical activity and radiographers 26 hours a week then 1 clinical minute for a radiologist is equivalent in value to 1.3 clinical minutes for a CSR (i.e. 34/26 = 1.3).

6.4 Data collection

The following describes the collection of data.

6.4.1 Cost data

The Business Manager for the X-ray department at Trust A provided the cost data, mainly using the questionnaire in Annex 6.

6.4.2 Time taken to report

Chapter 4 presents the findings of the study that assessed the effect of introducing CSR reporting A&E plain radiographs of the appendicular skeleton at Trust A. A retrospective sample of reports and radiographs was used so there was no record of the time it took CSRs and radiologists to report. The same sample of radiographs was used in the study presented in Chapter 5 that assessed CSR and radiologist performance when reporting A&E and GP plain radiographs. The following describes how data was collected when CSRs and radiologists reported these A&E plain radiographs to estimate the time it took them to report.

In normal practice, medical secretaries provide radiologists and CSRs with radiograph bags, the relevant radiographs and the X-ray request form. For both reporting groups, the secretaries are responsible for returning radiographs to the relevant source. It was not necessary to time this part of the reporting process, as it was the same for both groups. The remaining elements of the reporting process were divided into two tasks: the time it took the radiologists and the CSRs to report the radiographs and for the report to be recorded; the time it took to type reports.

Both CSRs and radiologists were provided with forms and a stopwatch to record the time, in seconds, that it took to report the sample of radiographs used for the studies presented in Chapters 4 and 5. The reference number on the top right-hand corner of the bags used to store the radiographs corresponded with the number on the proforma used to record a report. Both groups began timing themselves, having matched the reference number on the radiograph bag with the number on the proforma. The radiologists stopped timing, having recorded their reports onto tape. The CSRs stopped timing, having recorded the report onto the proforma and noting whether further advice was required. This process simulated normal clinical practice.

In normal practice, secretaries enter the radiologists' reports and the CSRs their own reports into the IT system. It was assumed that the time it took the radiologist to dictate the report onto tape was the time it took the secretary to type the report. A time sheet was designed for the CSR to record how long it took to type reports for different body areas during normal practice. The time recorded was from the point of identifying the relevant report to its entry into the computer system.

6.4.3 Number of radiographs reported

The Business Manager for the X-ray department at Trust A provided data on the number of plain radiograph X-ray examinations performed for different body areas and for A&E patients during the year 2001/2. Data was also provided on the number of examinations reported by CSRs and radiologists during that year.

6.5 Cost analysis of radiographer reporting A&E plain radiographs

The analysis was conducted from the perspective of the X-ray department at Trust A. The consequences of CSR and radiologist reporting were assumed to be the same so a cost-minimisation analysis was performed.

The GDP and HCHS indices were used to inflate the capital, training and staff costs to the year 2001/2 and they were then annuitized. A combination of the capital, training, and overhead costs and multiplication of the time (in clinical minutes) it takes CSRs and radiologists to report by their hourly rate were used to calculate the cost per radiograph reported. By multiplying the time CSRs and radiologists report by the number of radiographs reported it was possible to calculate the annual cost of CSR or radiologist reporting. Exact calculations are presented in Annex 6.

Average and incremental costs were also calculated. The average cost per unit of output is the total cost divided by quantity. The incremental cost is the difference in cost between CSR and radiologist reporting.

Sensitivity analysis was performed to test the impact of changes in assumptions about how certain costs should be included in the analyses. Several variables were identified as being suitable for inclusion, making the type of sensitivity analysis performed a scenario analysis. This approach involved constructing a series of scenarios that represented a multi-way analysis that recognises the uncertainty of various variables. Scenarios used were base, worse and best cases, presented in Annex 6.

6.6 Results

Results of the cost of CSR and radiologist reporting A&E plain radiographs of the appendicular skeleton at Trust A are presented here, followed by the results of the cost of extending the CSR role to include all A&E plain radiographs.

6.6.1 *The cost of reporting A&E plain radiographs of the appendicular skeleton*

Table 6.1 presents the costs associated with CSR and radiologist reporting inflated to the common year 2001/2. They include the setting up costs of CSRs reporting, such as the capital costs of purchasing new equipment and altering a room so the CSRs could report, and the costs of training CSRs. Staff costs and the overhead costs of the X-ray department also appear.

Table 6.1 Costs per year of plain radiograph reporting

Cost	Year	Item	Cost (£)	2001/2 (£)
Capital	1994/5	Computer terminal and bar code reader	405.38	488.47
	1994/5	Thermal Printer	988.18	1190.72
	1994/5	Worktops	750.00	903.72
	1994/5	Installation of PC and Printer	176.25	212.37
	1994/5	Installation of viewing boxes	495.00	596.46
	1994/5	Partition	965.00	1162.79
	Total		3779.81	4554.53
Training	1993/4	MSc Course – Registration	6000.00	7323.76
	1993/4	MSc Course - Travel expenses	426.42	520.50
	1994/5	MSc Course - Travel expenses	63.40	76.39
	1995/6	MSc Course - Travel expenses	14.95	17.51
	Total		6504.77	7938.16
Staff	1995/6	Basic radiographer (42 weeks pa, 35 hours pw)[a]	12376.00	16367.43
	1995/6	CSR (42 weeks pa, 35 hours pw)[b]	21002.80	27776.49
	1995/6	Radiologist - MC21 (42 weeks pa, 38 hours pw)[c]	40620.00	53720.51
	1995/6	Secretary – C3 (42 weeks pa, 35 hours pw)[d]	12788.00	16912.31
Overheads	2001/2	CSR reporting room ($10.28m^2$)		3240.85
		Radiologist offices ($88.64m^2$)		27944.46
		Radiologist hot reporting room ($22.3m^2$)		7030.25
		Secretary office ($32m^2$)		10088.25

a £11.13/hr
b £18.90/hr
c £33.66/hr
d £11.50/hr

Capital and training costs in Table 6.1 were annuitized to produce the equivalent annual cost (EAC). The annuity factor was selected from Table 2 page 94 in Drummond et al (1997). This approach assumes that annuity was paid in advance at the beginning of each year rather than in arrears. The UK Government recommended discount rate for public project including capital spending on the NHS is 6%. The rate used therefore varied from 5 per cent to 7 per cent for the purpose of the sensitivity analysis. Life of the capital costs was assumed to be five years for equipment like chairs, computer, printer and fifteen years for the partition used to provide a room for CSRs to report.

It was uncertain over how many years the training costs should be annuitized. Personal communication with the Society and CoR and the British Institute of Radiology suggested there was no reason why most radiographers should not work until retirement age. Furthermore, most radiographers have several years experience before being trained in image interpretation. Twenty-two radiographers who have received training to report at Bradford University were already qualified for a mean number of eleven years. It is unlikely that experienced radiographers who receive additional training will then leave the profession, so it seemed reasonable to annuitize the training costs over 15 years. The figure was changed from ten to twenty years for the sensitivity analysis.

To calculate the overhead costs (e.g. heat, light) presented in Table 6.1, the floor space for the different rooms used for reporting and the X-ray department in total ($3230m^2$) was identified. The overhead cost for the X-ray department during the year 2001/2 was £1,018,283. The percentage of floor space for the different rooms of the X-ray department was used to calculate the percentage of overhead costs that should be attributed to each room, but these are also shared costs. For example, the room that the CSRs use to report is sometimes used for meetings, and by radiologists to report, so the capital costs incurred in setting up the CSRs reporting room are also shared costs. Assumptions made about the allocation of the overhead and capital costs to CSR or radiologist reporting were changed for the sensitivity analysis.

The other important data to consider was the time it takes the two professions to report and the number of radiographs reported. Table 6.2 presents the average time in seconds that CSRs and radiologists report A&E plain radiographs. For the radiographs that the CSRs currently report, those of the appendicular skeleton, the time it took the CSRs to interpret and record a report (Time A – 47 seconds) was similar to the time it took consultant radiologists to dictate a report to tape (Time A – 43 seconds). However, CSRs spent additional time entering a report into the computer system (Time B – 19 seconds), which the secretaries did for the radiologists. In contrast, the CSRs took approximately twice as long to interpret and record a report (Time A – 103 seconds) for the remaining body areas, whereas for the radiologists it only took them a little longer than for the appendicular skeleton (Time A – 57 seconds). It was assumed that the time spent by the CSRs entering the reports into the computer system for the remaining body areas (Time B) was the same it took them to do this for an appendicular skeleton report i.e. 18.75 seconds.

Table 6.2 *CSR and radiologist A&E radiograph reporting times (seconds)*

Body Area	**Radiologist**	**CSR**	**CSR**	**CSR**
	Time A	Time A	Time B	Time A + B
Upper Limb	42.61	48.03	20.8	68.83
Lower Limb	42.39	46.23	16.8	63.03
Sub-total	42.50	47.13	18.75	65.88
Pelvis	53.45	109.42		
Head	50.13	63.63		
Spine	79.47	112.00		
CTA	43.73	125.56		
Sub-total	56.70	102.65		
TOTAL	51.96	84.15	18.75	102.90

These costs, and the time (in clinical minutes) it took a CSR or radiologist to report, were used to calculate the cost per radiograph reported. For the purpose of the sensitivity analysis it was possible to vary the CSR clinical minute relative to the radiologist clinical minute from 1.2 to 1.4.

To calculate the annual cost of CSR and radiologist reporting, the number of radiographs performed during the year 2001/2 was multiplied by this estimate. There were 9713 A&E plain radiographs of the appendicular skeleton.

Table 6.3 presents the worst, best and base case scenarios of the cost of CSR or radiologist reporting A&E radiographs of the appendicular skeleton at Trust A. Assumptions made for each scenario and exact calculations are presented in Annex 6. For the worst case scenario, CSR reporting would cost the X-ray department £4524 per annum, but this reflects the cost of CSR reporting when several assumptions are made in favour of radiologist reporting. The best case scenario shows that CSR reporting should save the X-ray department £4528 per annum, assuming that the CSRs take the same time to report as radiologists and a CSR and radiologist clinical minute are equivalent. For the base case analysis, which reflects the most realistic set of assumptions, CSR reporting should save the X-ray department £361 per annum, indicating no obvious cost savings or losses from CSRs reporting A&E radiographs of the appendicular skeleton at Trust A. However, if secretaries were to type CSR reports, this average cost would be reduced to £0.66/radiograph reported.

Table 6.3 *Cost of reporting A&E appendicular skeleton radiographs*

Scenario	Reporting Policy	Annual Total Cost (£)	Average Cost (£/radiograph reported)
Worse	Radiologist	5425.02	0.54
	CSR	9948.91	1.02
	Increment	- 4523.89	- 0.48
Best	Radiologist	9414.91	0.97
	CSR	4886.78	0.50
	Increment	4528.13	0.47
Base	Radiologist	7498.17	0.77
	CSR	7137.03	0.73[a]
	Increment	361.14	0.04

[a] £0.66 if a secretary types CSR reports

6.6.2 The cost of reporting all A&E plain radiographs

Findings presented in Chapters 2 and 5 support the potential for extending CSRs' reporting role at Trust A to include the remaining body areas for A&E patients. Based on this evidence and audits performed at Trust A, the X-ray department were considering extending the CSRs' role in this way. However, Table 6.2 showed the average time it takes the CSR to interpret and record their report for an A&E radiograph (84 seconds) was considerably longer than for an A&E radiograph of the appendicular skeleton (47 seconds). In addition, they spent 19 seconds entering a report into the computer system. This may increase the cost of CSR reporting compared with radiologists. Annex 6 presents the assumptions made and the calculation of the cost of CSR and radiologist reporting A&E plain radiographs.

Table 6.4 Cost of CSR and radiologist reporting all A&E radiographs

Scenario	Reporting Policy	Annual Total Cost (£)	Average Cost (£/radiograph reported)
Appendicular skeleton	Radiologist	7498.17	0.77
	CSR	7137.03	0.73 [a]
	Increment	361.14	0.04
All body areas	Radiologist	14853.15	0.74
	CSR	17763.73	0.89 [b]
	Increment	- 2910.58	- 0.15

[a] £0.66 if secretary types CSR reports

[b] £0.82 if secretary types CSR reports

During the year 2001/2 10,660 radiographs were performed for the remaining body areas. The Business Manager at Trust A was of the opinion that all these radiographs are reported, so around 20,000 A&E radiographs for all body areas were reported that year. Table 6.4 presents the base case cost of CSR and radiologist reporting A&E radiographs of the appendicular skeleton and all body areas, showing a potential for cost savings (£361) when the CSR report A&E plain radiographs of the appendicular skeleton but cost losses (£2911) when they report all A&E plain radiographs. This deficit may diminish as the CSRs gain more experience reporting A&E radiographs for the body areas they do not currently report, in turn resulting in them reporting more quickly. Table 6.2 showed that was feasible, as when reporting A&E radiographs of the appendicular skeleton, the time it took the CSRs to interpret and record their reports (47 seconds) was comparable with the time it took consultant radiologists to dictate their reports to tape (43 seconds). Further cost savings could be made if secretaries were to type the CSR reports, as secretarial resources are, in general, less expensive than radiographers.

6.7 Discussion

When Piper et al (2000) explored the implications of selectively trained radiographers rather than radiologists reporting A&E radiographs in four NHS Trusts, the additional costs ranged from nil to £15,000 per annum. The cost of CSRs at Trust A reporting A&E radiographs of the appendicular skeleton per annum range from a saving of £4528 to a loss of £4524. With a base case saving of only £361 per annum, the findings are similar to Piper et al (2000) in that no obvious cost savings or losses were found. However, a saving of £361 is a negligible amount of money that could have been used in some other way to achieve a health outcome.

Lack of cost savings was a result of training the CSR, modifying rooms, and providing them with the equipment they need, which amounts to a few thousand pounds even after annuitisation over several years. Extra costs are associated with the time it takes the CSR to report radiographs and the value of a CSR clinical minute relative to that of a radiologist. It is only when reporting times and clinical minutes are assumed to be the same for each profession, as for the best case scenario, that there is the potential for cost savings. As CSRs acquire the same experience as consultant radiologists, it is feasible that they could spend the same time reporting. Furthermore, if secretaries were to type the reports that CSRs dictate to tape, as they do for radiologists, this could result in further cost savings.

Introducing CSR reporting may also have resource management implications. Who will perform the duties that they used to do and how will radiologists spend the time freed from not having to report as many radiographs? To explore how this skill mix initiative might result in CSRs' and radiologists' time being used more productively, a survey was conducted to assess whether different professionals at Trust A found CSR reporting A&E plain radiographs of the appendicular skeleton to be acceptable (Brealey et al, 2002). A self-answer questionnaire, using the Likert scale, measured the attitude of different professionals at the hospital (i.e. CSR, radiologists, A&E consultants). An open-ended section investigated how CSR reporting affected their workload and freed radiologists to perform other tasks.

The self-answer questionnaire found that, except for one radiologist who was uncertain, the different professionals agreed that CSR reporting A&E radiographs of the appendicular skeleton was acceptable. In the open-ended section, the CSRs acknowledged that reporting reduced their involvement in general Radiography and they no longer worked in A&E or speciality areas (e.g. Computed Tomography, Angiography). They also commented that reporting radiographs takes '25 to 50 per cent of my time'. In contrast, the radiologists thought that CSR reporting frees only a little of their time, possibly 'about ¾ hour per week'. The fact that this task was shared between eight radiologists, as opposed to two CSRs, might explain why only a little of the radiologists' time appears to have been freed.

These qualitative findings suggested that whereas the reporting of radiographs had a dramatic impact on activities performed by the CSRs, there was a negligible change to the radiologists' work. Although the survey provided some evidence about the resource implications of introducing CSR reporting, it would have been more accurate to prospectively monitor the affect of introducing CSR reporting on changes in staff workload and changes in the time allocated to different tasks in the X-ray department.

Another consequence of introducing CSR reporting at Trust A with possible resource implications concerns the number or proportion of radiographs reported. Indeed, an initial reason for the scheme was to ensure that all radiographs could continue to be reported. Whether this could be achieved is both interesting and important because of the health and cost trade-off between all or only some radiographs being reported. For example, if all radiographs are reported there is the time and cost associated with them being reported and subsequent resource use when managing patients. If not all radiographs are reported, this saves time and money both of the report and the unnecessary follow-up of patients due to an error in the interpretation of the radiograph. Resources would also be wasted in producing the radiograph and the cost implications of missing some abnormalities. Prospective collection of data was required to assess whether A&E and other radiographic examinations at Trust A continued to be reported at the same level of activity.

This study was also performed at a single hospital in quite a rural area where CSRs were compared with consultant radiologists. The findings might not be generalisable to a teaching hospital, in an urban area, where radiologists of varying seniority are involved in reporting A&E radiographs, and there is variation in the number of hours that radiologists work and are clinically productive.

Nevertheless in terms of assessing the effect of CSR reporting A&E plain radiographs on the efficiency of the X-ray department at Trust A, the associated costs were comprehensively identified, measured and valued. A scenario analysis was also conducted, acknowledging uncertainty in key assumptions about, for example, the time period over which to annuitize the training costs, the allocation of overhead costs, and the value of a clinical minute. Performing the study from a narrow perspective does seem to be a valid assumption, based on the evidence presented in the preceding chapters of the thesis. It is also unlikely that the occasional missed subtle fracture will have huge implications to NHS service provision, patient quality of life and cost savings. Whilst the prospective collection of data on changes in staff workload would be useful to reflect the broader resource management implications of CSR reporting, this is not crucial to the cost analysis as an assessment of service efficiency.

6.8 Conclusion

With a base case saving of £361 per annum, there do not appear to be any obvious cost savings or losses to the X-ray Department at Trust A from substituting CSRs for radiologists when reporting A&E plain radiographs of the appendicular skeleton. For CSR reporting of all A&E radiographs, there were potential cost losses of £2911 per annum, which may diminish as the CSRs gain more experience reporting A&E radiographs for the body areas they do not currently report, and in turn reporting them more quickly. Further cost savings could be made if secretaries were to type the CSR reports, as secretarial resources are, in general, less expensive than radiographers.

The next, and final, chapter will use the evidence presented in this thesis to discuss the conclusions that can be drawn about selectively trained radiographers reporting plain radiographs, and the necessity for future research.

Chapter 7

Conclusions and future research opportunities

7.1 Introduction

The aim of this thesis was to determine whether selectively trained radiographers should report plain radiograph X-ray examinations.

The background to radiographer reporting and the complexities of measuring reporting performance were discussed in the first chapter and results of a systematic review synthesising the existing evidence on the accuracy of radiographer plain radiograph reporting were presented in Chapter 2. The methods for measuring reporting performance were developed and assessed in Chapter 3, so that valid and reliable data would be produced by the primary studies of the thesis. Chapters 4 to 6 present the results of the primary research in terms of accuracy, clinical effects and associated costs of clinical specialist radiographers (CSRs) and consultant radiologists reporting plain radiographs at Trust A. The current chapter seeks to use the evidence presented in this thesis and related literature to discuss the conclusions to be drawn and the future research necessary. It concludes with a final statement on the viability of selectively trained radiographers reporting plain radiographs.

7.2 Discussion about radiographer plain radiograph reporting

A synopsis of the current opportunities and threats to radiographer reporting as a skill mix initiative is presented to facilitate subsequent discussion on the results of the thesis and other related research when addressing whether selectively trained radiographers should report plain radiographs. The section ends with a summary of recommendations for clinical practice and research.

7.2.1 Opportunities and threats to radiographer reporting

The Government maintains that one of its priorities is the modernisation of the NHS. The *NHS Plan* (DH, 2000) and *Meeting the Challenge* (DH, 2000) describe how modernising pay structures, increasing earnings and improving staff working lives in the form of advanced practitioner and consultant posts will achieve this goal. The new posts are to provide opportunities for career development through greater innovation in the deployment of staff. Detailed changes to the NHS pay system are outlined in

Agenda for Change (DH, 2003) and *Knowledge & Skills framework* (DH, 2003), underpinning the government commitment to invest and reward NHS staff with 'equal pay for work of equal value'. Government rationale for the initiatives is to improve recruitment, retention and staff skills and morale and in turn help it meet other targets by treating more patients, more quickly and delivering a higher quality of care. The implementation of more effective and efficient services should, in turn, help improve outcomes for patients.

Growth in skills in the profession of Radiography is an ongoing need, as is the search to retain professional identity and status in a dynamic health care climate (O'Connor, 1996). The recent government initiatives provide radiographers with the opportunity to ensure their continuing development is successful and enduring. In order for them to fit the criteria for the newly established posts, they must demonstrate expertise and leadership in the research, clinical and professional environment. Opportunities in the form of skill mix initiatives are fundamental in providing radiographers with the education and training required to maximise and improve their skills and knowledge. Skill mix can help develop radiographers' research skills in a specialist clinical field, just as working in an extended role can provide them with the experience to bring innovation, clinical leadership and strategic direction within the NHS and the ability to integrate research evidence into practice.

Despite the advantages of skill mix to radiographers and the benefits to radiologists as a consequence of their time being freed to perform more specialist and complex investigations, direct substitution could de-skill radiologists for the tasks they no longer perform and thereby threaten standards (RCR, 1999). Furthermore, whilst the new consultant posts mean that allied health professionals can earn more, the result could mean more costly skill mix initiatives. The danger is, then, that the same incentive for improving staff recruitment, retention and morale could threaten the justification for introducing skill mix.

Tables 7.1 and 7.2 present three-month vacancy figures from the Department of Health annual censuses of the NHS workforce for consultant and allied health professionals. Compared with all consultants, there is a higher percentage of vacancies for radiologist posts (Table 7.1). The data also suggests that during the period 1999 to 2003 the gap in vacancy rates between radiologists and all consultants widened. Table 7.2 shows that whilst vacancy rates for diagnostic radiographer posts are comparable to all allied health professionals, each year has seen an increase in the percentage of posts vacant. Notably, for the year 2003 there was an increase in vacancy rates between diagnostic radiographers and all allied health professionals.

Table 7.1 Vacancy rates (%) for medical posts from 1999 to 2003

Year	All consultants	Radiology	% Difference
2003	1264 (4.7%)	139 (7.6%)	2.9%
2002	950 (3.8%)	140 (8.0%)	4.2%
2001	670 (3.0%)	70 (4.6%)	1.6%
2000	600 (2.8%)	70 (4.8%)	2.0%
1999	470 (2.3%)	50 (3.1%)	0.8%

Table 7.2 Vacancy rates (%) for allied professional posts from 1999 to 2003

Year	Allied Health Professionals	Diagnostic Radiographers	% Difference
2003	2176 (4.8%)	599 (6.1%)	1.3%
2002	2190 (5.0%)	530 (5.5%)	0.5%
2001	1820 (4.3%)	420 (4.4%)	0.1%
2000	- (3.6%)	- (4.1%)	0.5%
1999	- (2.1%)	n/a[a]	n/a[a]

[a] Before 2000 diagnostic and therapeutic radiographer figures were combined

National surveys illustrate an increase in the number of radiographers reporting A&E radiographs in recent years from four Trusts (Paterson, 1995) to 37 (Price et al, 1999). Although it could be argued that vacancy rates would be even worse without such initiatives, their introduction does not seem to be reflected in improved retention and recruitment. The national shortage of radiologists and radiographers also results in delegation being compromised by the lack of healthcare professionals willing, capable and competent to undertake the delegated tasks, further compromising their ability to use existing skill mix initiatives, let alone develop new ones. All this threatens the delivery of the *NHS Plan* (DH, 2000) and other government initiatives. Given a shortage of staff, how can quality and the immediacy in delivery of healthcare be improved?

The following section discusses the findings of this thesis and other research in relation to the role of selectively trained radiographers reporting plain radiographs whilst taking into consideration the opportunities and threats of this skill mix.

7.2.2 *Radiographer plain radiograph reporting*

7.2.2.1 A&E plain radiograph reporting

A&E departments are integral to the health service in the provision of front-line diagnosis for patients in the NHS. Hundreds of thousands of patients pass through A&E departments each year and are usually referred to the X-ray department for a plain radiograph examination. During the year 2001/2, for example, there were 20,373 A&E plain radiograph examinations performed at Trust A.

Historically, the reporting of radiographs has been the domain of radiologists. Not until the early 1970s was it first proposed that a radiographer could identify 'normal' and 'abnormal' radiographs without prolonged, complex training (Swinburne, 1971). Over a decade later Berman et al (1985a) assessed the introduction of a scheme whereby radiographers marked A&E radiographs with a red dot to alert casualty officers to the possible presence of abnormalities. Renwick et al (1991) took this a step further, with a scheme that involved radiographers with no formal training in pattern recognition interpreting A&E radiographs in a triage role, classifying radiographs as normal, insignificantly abnormal, or significantly abnormal. Research in the mid 1990s by Loughran (1994) and Robinson (1996a) assessed the ability of selectively trained radiographers to make written reports of A&E plain radiographs. And more recently, A&E nurses have been assessed in an A&E radiograph reporting role (Meek et al, 1998). All this has resulted in various professions (e.g. casualty officers, radiologists, radiographers, nurses) and a variety of methods (e.g. red dot, triage, reporting, hot/cold systems) being used for commenting on A&E plain radiographs.

Although research is often conducted to address heterogeneity in clinical practice, it could be argued that variation in who and how A&E radiographs are commented upon is important. Viewing these radiographs is useful for developing the image interpretation skills of casualty officers and junior radiologists, and as an opportunity to extend and underpin radiographers' and A&E nurses' professional roles. Government policy focuses not only on patient needs, but also on those of the staff in the NHS. The various ways in which A&E plain radiographs are interpreted is a prime example of flexible team working between different clinical professions. A combination of radiographers in a red dot system, casualty officers in a triage role, and reporting by radiologists or selectively trained radiographers has the potential for reducing errors.

Research presented in Chapter 2 provides evidence of varying quality, from several studies performed in different settings, that given some training, radiographers report A&E radiographs to a high level of accuracy. In addition, there is no evidence of a significant difference in sensitivity and specificity between selectively trained radiographers and radiologists of varying seniority reporting A&E radiographs. The primary research presented in Chapters 4 to 6 provides further evidence that selectively trained radiographers report A&E radiographs with no significant difference from consultant radiologists. Neither is there any clear evidence of a significant difference in the clinical effects of CSR and radiologist reports, nor any obvious cost savings or losses from substituting CSR with radiologists. In summary, all the evidence indicates that selectively trained radiographers can report A&E radiographs accurately and without a detrimental effect on patient care, outcome and associated costs.

Further evidence is provided by the systematic review, that radiographers red dot or triage A&E plain radiographs to a high level of accuracy and that the combined performance of radiographers and A&E staff can improve accuracy, although the low quality of the studies suggests that more rigorously designed research is required. In addition, the review was designed to address the accuracy with which radiographers report plain radiographs. Research is now needed to explicitly address the accuracy, effectiveness and costs of the different methods of reading A&E radiographs such as red dot, triage and reporting by different health care professionals.

Radiographs must not only be interpreted *accurately* but also *promptly*, as emphasised in several college recommendations (RCR, 1993; RCR, 1995; ACR, 1995). The public is greatly concerned about the time it waits for treatment, for example in casualty departments (DH, 2000). In response, the government target is that by 2004, waiting time in A&E should be no longer than four hours from arrival to admission, transfer or discharge (DH, 2000). Historically, surveys have shown that only 88 per cent (James et al, 1991), 83 per cent (Berman et al, 1985b), and 80 per cent (Beggs & Davidson, 1990) of Radiology departments report all A&E examinations, and indeed that 33 per cent of radiologists *never* report 10 per cent or more A&E radiographs (Rose & Gallivan, 1991). 'Hot' (or immediate) reporting services are only available in 2.6 per cent of departments and 49 per cent do not have A&E radiographs reported under 48 hours (James et al, 1991). An Audit Commission (1995) survey drew attention to the dissatisfaction of some clinicians, including those from A&E, that reports were not provided for all examinations and that a significant percentage were not received in time to influence patient management. Could the introduction of selectively trained radiographers reporting A&E plain radiographs potentially address this problem?

A recent study evaluating the implementation of radiographers reporting A&E skeletal radiographs in four NHS Trusts (Piper et al, 2000) addressed this question. At two Trusts, a significant increase was noted in the percentage of A&E skeletal radiographs reported after the implementation of radiographer reporting. Similarly, the percentage of reports available within five days of the examination being performed saw a significant increase at two Trusts. These findings support the potential of this skill mix to allow X-ray departments meet college quality standards, though the study had several limitations. First, the data was only collected during a three or four-month period before and after radiographer reporting was implemented. At some Trusts there was also a gap of several months between the collection of data before radiographer reporting was implemented and when it was eventually implemented: collecting data at different times of the year could explain study findings. Second, the data was aggregated so that the chi-square test for significance could be used, but the data should have been analysed as it was collected (i.e. weekly intervals), using more sophisticated analyses to adjust for seasonal variation. Third, there was no control group to eliminate other threats that could explain these changes. More rigorous research, or audit, is recommended to further assess whether radiographer reporting can achieve this potential, and to establish whether timeliness of reports and all radiographs being reported does bring about the expected benefits to patient care and outcome.

Finally, Chapter 6 demonstrated that there were no obvious cost savings or losses from introducing selectively trained radiographers reporting A&E plain radiographs of the appendicular skeleton. This research provided similar findings to Piper et al (2000) but there are problems in generalisability from a single district hospital to different hospital settings.

7.2.2.2 GP plain radiograph reporting

The RCR also recommend that X-ray departments provide GPs with a prompt reporting service (RCR, 1996), with 100 per cent of GP examinations issued with a report within 24 hours of the arrival of the patient (RCR, 1995b) as the standard. Delays in the availability of a report to GPs may affect patient outcome and lead to an increase in other, more expensive, investigations. A national audit of 102 hospitals in the UK recently investigated X-ray departments' adherence to this standard, and found that only 50 per cent of GP examinations were reported after 22 hours and 95 per cent were reported after 120 hours. The audit concluded that X-ray departments are often unable to meet targets set for providing GPs with a prompt reporting service (RCR, 2000). So it is possible that selectively trained radiographers reporting GP radiographs could help X-ray departments meet these standards.

The results presented from the feasibility study in Chapter 5 demonstrated that there was no significant difference in the efficacy of CSRs or radiologists reporting GP plain radiographs. This would indicate that CSRs could be introduced to the rota for reporting GP plain radiographs without a detrimental effect on report quality. A pragmatic, appropriately powered study is now recommended to assess the accuracy, clinical effects and costs of CSR reporting GP plain radiographs during clinical practice and whether it can improve the timeliness of reports.

The studies presented in Chapters 4 and 5 also demonstrate that CSR and consultant radiologists report GP plain radiographs significantly less accurately than A&E plain radiographs. In addition, whereas around 20 per cent of incorrect reports of A&E radiographs may result in a clinically important effect on patient management, the figure is only around 50 per cent for incorrect reports of GP radiographs (Chapter 4). Findings in Chapter 5 suggested that changes in GPs' confidence in their diagnosis and management plans, based on incorrect CSR and radiologist reports, was significantly lower than for the A&E clinician. The lower level of accuracy with which GP radiographs were reported compared with A&E radiographs might be explained by the more complex radiographic appearances, which in turn might influence the confidence of the reporting CSR and radiologist and, subsequently, the confidence of the referring clinician. However, if GP confidence in diagnosis and management planning is low and does not change after reading the report, the appropriateness of the initial referral is brought into question.

Therefore, another method of reducing radiologist and radiographer workload and expediting the availability of reports is to improve the quality of GP referrals. Radiography of the lumbar spine, for example, is a common referral from primary care. During the year 2001/2 there were 21,247 GP plain radiograph examinations performed at Trust A of which 1,857 were of the lumbar spine. In the same year, Kendrick et al (2001) published a randomised trial, which concluded that patients with low back pain, referred from primary care for radiography of the lumbar spine, was not associated with improved clinical outcomes. Another randomised trial provided evidence that the routine attachment of educational messages to radiographs can significantly reduce GP referrals for lumbar spine radiographs (Eccles et al, 2001). The challenge is to implement these findings in practice, particularly with the continuing increase in the number of plain radiographs being performed in X-ray departments (RCR, 2002).

7.2.3 Recommendations about radiographer plain radiograph reporting

Evidence indicates that selectively trained radiographers and radiologists report A&E plain radiographs to a comparable level of accuracy and effect on patient diagnosis, management, outcome and the associated costs. X-ray departments should invest in selectively trained radiographers reporting A&E radiographs, if this can help meet local demands.

Radiographers can also red dot and triage A&E plain radiographs to a high level of accuracy. This can be improved in combination with other professionals, although the evidence for this was of low quality. Further research is recommended to more rigorously assess the different methods of reading A&E plain radiographs.

There was also evidence that selectively trained radiographers could report GP plain radiographs with no detriment to accuracy and the subsequent clinical effects. Before radiographer reporting GP plain radiographs can be recommended, further research is needed, conducted during clinical practice at multiple sites and including an assessment of accuracy, clinical effects and costs, and the timeliness of report availability.

Finally, whilst college recommendations for the prompt reporting of all examinations are admirable, it should be established whether timeliness of reports and all radiographs being reported *does* bring about the expected benefits to patient care and outcome at reduced costs.

7.3 Methodological issues about radiographer reporting

When the background research for the thesis was conducted, it became apparent that studies of radiographer plain radiograph reading performance were of varying quality, based on methodological papers as to how studies should be designed to evaluate diagnostic tests. There was also an absence of evidence concerning clinical effects of radiographer plain radiograph reporting performance in terms of, for example, the effect on the referring clinicians diagnosis, choice of patient management, patient outcome and the associated costs. Supplementary objectives were subsequently included in the systematic review about threats to the validity of studies that assess radiographer plain radiograph reading performance and the subsequent clinical effects and associated costs. Findings from the systematic review helped to inform the design of the primary studies presented in Chapters 4 to 6. Some of the salient methodological issues that arose when performing the systematic review and primary studies presented in this thesis but as yet not discussed are addressed here.

7.3.1 Development of a quality criteria checklist

Recent research into improving the evaluation of diagnostic tests (Deeks, 2001; Deville et al, 2002; Lijmer et al, 1999) has been considerable. In particular, the Standards for Reporting of Diagnostic Accuracy (STARD) initiative, NHS Centre for Reviews & Dissemination and Cochrane Collaboration are developing checklists to appraise the quality of studies of diagnostic accuracy. However, at the time the systematic review presented in Chapter 2 was designed, there was no clear guidance, unlike for randomised trials, on appraising the quality of studies that assess the accuracy of diagnostic tests. So the systematic review sought to identify threats to the validity of studies assessing radiographer plain radiograph reading performance resulting in the development of the quality criteria checklist component of the data extraction form (discussed in Chapter 2).

When designing the checklist it was only feasible for the author and radiographer reviewer to identify threats to study validity that could be used in appraising the quality of the studies included in the review. An alternative approach is to conduct a literature search identifying an exhaustive list of threats to study validity and then seek expert judgment of the importance of the criteria. Such agreement by experts about which criteria to include, would enable these to be subjectively weighted for importance and the checklist tested for intra/inter-rater reliability. Another option is to conduct a systematic review about all health care professionals reporting plain radiographs and using the extension of Littenberg & Moses' regression method (described in Chapter 2) empirically determine the quantitative effect of study design limitations on reporting accuracy studies. Thereby, the weight of importance attached to individual criteria could be determined.

The quality criteria checklist that was developed and discussed in Chapter 2 was found to be reliable and useful for identifying all sources of threats to the validity of studies included in the review and the prevalence with which these threats were present. Further research would be useful to develop a more concise, reliable and valid checklist for assessing the quality of radiograph reading accuracy studies to facilitate their future conduct, design and critical appraisal.

7.3.2 Choice of study design

Design of primary studies to assess the effect of introducing a skill mix initiative like selectively trained radiographers reporting plain radiographs requires several methodological issues to be addressed.

A randomised trial might, for example, be designed to determine whether selectively trained radiographers should report GP plain radiographs or whether they should continue to be reported by radiologists. Randomisation would eliminate bias in the selection of radiographs to be reported by the radiographers and radiologists and provide a basis for statistical analysis. It would also permit the prospective collection of data on the subsequent effects of reports on patient management and outcome, and the associated costs. The methodological issue of interest and in need of further discussion is the calculation of the sample size for addressing the research question.

One of the limitations of the primary research (as discussed in Chapters 4 and 5) was the low power to detect a statistical and clinical important difference. A study with a small sample size and subsequently low power can result in failure to reject a false null hypothesis, i.e. declaring that a difference does not exist when in fact it does (a type II error). The study presented in Chapter 4, for example, was designed to test whether the introduction of CSR reporting A&E radiographs resulted in a significantly inferior or superior quality service. However, the usual impetus for skill mix initiatives is that the intervention does not necessarily bring about significant benefits to the quality of the service, but ensures the same quality service at less cost. Arguably, the aim was not to test whether the CSR produced a clinically significant difference in A&E radiograph reporting performance but to whether the reporting performance after their introduction was equivalent to the radiologists, or of no (or little) clinical significance. This is a particularly salient element, as it will determine the power of the study and the number of patients to be recruited and consequently the resources required to conduct the study.

7.3.3 *Assessment of resource management implications*

A final methodological issue for consideration is the resource management implications of introducing a skill mix initiative such as selectively trained radiographers reporting plain radiographs. Economic evaluations are concerned with the costs and benefits of introducing an intervention, compared with usual practice, assuming that everything else remains constant. But a skill mix initiative, like radiographer reporting, is often introduced so that the substitution of the doctors by the non-medical profession frees up the doctors, allowing for more productive management of time and skills of both professions. Integral to the introduction of a skill mix, is the desire to improve the creative and flexible use of all personnel resources within a dynamic X-ray Department and not for everything else to stay constant.

For example, basic radiographers substitute the role of senior radiographers who are now reporting radiographs. The development of the radiographers in a reporting role may also benefit the other radiographers in the department, if it improves their understanding of the production and interpretation of radiographic images. The radiologists also benefit in that they no longer perform a task they perceive as a chore and can concentrate on conducting specialist investigations. The skill mix, then, should help improve the management of resources within the X-ray department as staff are being used to fulfil their potential. Job satisfaction should increase as a result, as should the recruitment and retention of staff because of the more diverse career opportunities.

It was beyond the scope of the economic evaluation presented in Chapter 6 to identify, measure and value all the costs and benefits of introducing a skill mix initiative like radiographer reporting. Future economic evaluations should prospectively monitor the ramifications of its introduction to the wider activities within the X-ray Department and whether this results in improved job satisfaction. The impact of the skill mix initiative from the perspective of the staff should be reflected, rather than focusing on the consequences to the patient in terms of diagnosis, management and outcome.

7.4 Conclusion

The subject of non-medically qualified staff reporting radiographs has been debated and contested almost since the discovery of X-rays by Röntgen in 1895. When policy changed with the NHS and Community Care Act in 1990 to meet the increasing demand in health care, this provided X-ray departments with the opportunity to introduce selectively trained radiographers as a skill mix initiative to clinical practice.

The results of this thesis provide evidence that they can report A&E plain radiographs to a comparable level of accuracy to radiologists and with no detrimental effect on patient diagnosis, management, outcome and the associated costs. This evidence underpins the increasing introduction of this skill mix in an X-ray Department, when it can help meet local demands. Further research is needed before the same conclusion can be drawn about selectively trained radiographers reporting GP plain radiographs.

The new government initiatives will help secure the introduction of skill mix in the form of consultant posts for allied health professionals. Whilst more creative and flexible team working between staff of different professions is desirable, this does not guarantee the delivery of more effective and efficient services. Not only should the needs of staff and patients be considered but also the prudent use of inevitably scarce resources. The challenge to X-ray departments and the rest of the NHS is to use the evidence to inform local decisions about when the creative use of different professionals is an efficient use of resources.

Annexes

Annex 2.1 Search strategies and references

Search Strategies

Detailed search strategies are provided for all the electronic databases searched. Table A2.1 explains the abbreviations and commands used.

Table A2.1 Definition of commands and abbreviations

Abbreviation or command	Definition
*	Truncation symbol for MEDLINE, BIDS, EMBASE, CINAHL, PsycLIT, NRR, The Cochrane Library
Mesh	Indexing term in MEDLINE
de	Indexing term in CINAHL and EMBASE
?	Wildcard symbol in MEDLINE; truncation symbol in SIGLE
*.me	Indexing term in NRR and The Cochrane Library

Search strategies to check for ongoing and existing reviews

MEDLINE (1966-9/97)

No.	Records	Request
1	170764	REVIEW
2	576355	REVIEW in PT
3	74933	REVIEW in TI
4	344	RADIOGRAPHER?
5	670318	#1 or #2 or #3
6	21	#4 and #5

DARE (no temporal restrictions)

No.	Records	Request
1	3260	REVIEW
2	4	RADIOGRAPHER*
3	0	#1 and #2

Electronic database search strategies

MEDLINE via the Ovid Web Gateway (1971 to end of October 2002)

No	Records	Request
1	19823	diagnostic-errors
2	101013	sensitivity-and-specificity
3	120121	#1 or #2
4	337612	explode "Radiography"/ all subheadings
5	12724	explode "Radiology"/ all subheadings
6	26207	radiography
7	234872	radiog$ or radiol$
8	1035945	report$
9	4287	triage
10	5243	x ray film$
11	1460545	#4 or #5 or......#10
12	27990	#3 and #11
13	487	radiographer$
14	25	#12 and #13
15	25	limit 15 to yr >=1971

Science Citation Index-Expanded via ISI Web of Science Service (1981 to end of October 2002)

No	Records	Request
1	255	radiographer*

CINAHL via the Ovid Web Gateway (1982 to end of October 2002)

No	Records	Request
1	1091	diagnostic-errors
2	3540	sensitivity-and-specificity
3	4590	#1 or #2
4	8708	explode "Radiography"
5	441	explode "Radiology-Service"
6	1755	radiography
7	8207	radiog$ or radiol$
8	50871	report$
9	1529	triage
10	97	x ray film$
11	64446	#4 or #5 or.....#10
12	1049	#3 and #11
13	362	radiographer$
14	15	#12 and #13

EMBASE via the Ovid Web Gateway (1980 to end of October 2002)

No	Records	Request
1	3587	observer-variation
2	81573	error
3	753434	diagnosis
4	190041	injury
5	65391	diagnostic-accuracy
6	40681	diagnostic-imaging
7	1055450	#1 or #2...or #6
8	206416	explode "radiography"/ all subheadings
9	6425	explode "radiology"/ all subheadings
10	20985	radiography
11	158424	radiog$ or radiol$
12	882816	report$
13	1859	triage
14	4555	x ray film$
15	1141445	#8 or #9 or...#14
16	281614	#7 and #15
17	380	radiographer$
18	126	#16 and #17

NRR (to end of October 2002)

No	Records	Request
1	95	Radiographer*

The Cochrane Library – 2002 Issue 4 (to end of October 2002)

No	Records	Request
1	293	(Diagnostic and errors)
2	883	(Observer and variation)
3	1161	(#1 or #2)
4	65	Triage*.ME
5	5375	Radiography*.ME
6	101	Radiology*.ME
7	20918	Radio*
8	23001	(#4 or #5 or #6 or #7)
9	355	#3 and #8

PsycINFO via ISI Web of Science Service (1971 to end of October 2002)

No	Records	Request
1	41876	(pattern*) and (discrim* or strat* or identif* or perc* or recog*)
2	57293	(vis*) and (acuity or perc* or search or strat* or track*)
3	93806	explode "Cognitive-Processes" in DE
4	177272	#1 or #2 or #3
5	497	explode "Roentgenography" in DE
6	232	radiog*
7	691	#5 or #6
8	58	#4 and #7

SIGLE via WinSPIRS (1980 to end of October 2002)

No	Records	Request
1	23	Radiographer?

References of studies included

1. Remedios D, Ridley N, Taylor S, de Lacey G (1998). Trauma radiology: extending the Red Dot system. *British Journal of Radiology*, **71(S)**, 60.
2. Hughes H, Hughes K, Hamill R (1996). A study to evaluate the introduction of a pattern recognition technique for chest radiographs by radiographers. *Radiography*, **2**, 263-288.
3. Callaway MP, Kay CL, Whitehouse RW, Fields JM (1997). The reporting of A&E radiographs by radiologist trainees and radiographers. *RAD Magazine*, **23**, 31-32.
4. Balcam S, Hood A (1998). Are radiographers with suitable training competent to "cold" report plain films from the accident and emergency department? Personal Communication.
5. Robinson PJA (1996). Plain film reporting by radiographers - a feasibility study. *British Journal of Radiology*, **69**, 1171-4.
6. Berman L, de Lacey G, Twomey E, Twomey B, Welch T, Eban R (1985). Reducing errors in accident department; a simple method using radiographers. *British Medical Journal*, **290**, 421-2.
7. Renwick IGH, Butt WP, Steele B (1991). How well can radiographers triage X-ray films in the accident and emergency department? *British Medical Journal*, **302**, 568-69.
8. Eyres RD, Williams P (1997). Monitoring the effectiveness of a clinical reporting and educational programme: are radiographers getting it right? *British Journal of Radiology*, **70(S)**, 123.
9. Eyres RD, Williams P (1999). Results of a postgraduate training programme. Personal Communication.
10. Boynes S, Scally AJ, Webster AJ, Kay K (1999). Results of a postgraduate training programme. Personal Communication.
11. McMillan P, Paterson A, Piper K (1995). Radiographers' abnormality detection skills in radiographs of the chest and abdomen. *British Journal of Radiology*, **68(S)**, 21.
12. Manning D (1999). Results from audit process of students on a postgraduate course. Personal Communication.
13. McConnell JR, Webster AJ (2000). Improving radiographer highlighting of trauma films in the accident and emergency department with a short course of study – an evaluation. *British Journal of Radiology*, **73**, 608-12.

14. Henderson I (1999). Results of a postgraduate course. Personal Communication.
15. Raynor JS (1999). Audit study: 1995-1997. Grey Literature.
16. Loughran CF (1994). Reporting of fracture radiographs by radiographers: the impact of a training programme. *British Journal of Radiology*, **67**, 945-950.
17. Piper K, Paterson A (1997). Accuracy of Radiographers' Reports in Accident and Emergency Examinations of the Skeletal System. *European Radiology*, **7(S)**, 178-179.
18. Piper KJ, Godfrey RC, Paterson AM (2002). Reporting of non-A&E radiographic examinations by radiographers: a review of 6796 cases. *British Journal of Radiology*, **75 (S)**, 35.
19. Loughran CF, Raynor J, Quine FM (1996). Radiographer reporting of accident and emergency radiographs: a review of 5000 cases. *British Journal of Radiology*, **69(S)**, 128.
20. Timmis AJ, Burnett SJD (1995). Red Dot-Blue Dot: an alternative. *British Journal of Radiology*, **68(S)**, 41.
21. Ford L, Crawshaw I (1999). Results of audits performed at York District Hospital. Personal Communication.
22. Wilson J (1996). Extended role of the radiographer - plain film reporting project. *British Journal of Radiology*, **69(S)**, 24.
23. Seymour R, White P (1996). A & E reporting: relative performance of radiographers and radiologists. *British Journal of Radiology*, **69(S)**, 129.
24. Giles M (1989). Identification of abnormalities by the radiographer. HDCR Project Module F.
25. Bowman S (1991). Introducing an abnormality detection system by radiographers into an accident and emergency department: An investigation into radiographers' concerns about the introduction of such a system. *Research in Radiography*, **1**, 2-20.
26. Webster J, Gallagher D (1998). Radiography reporting in a community hospital. *British Journal of Radiology*, **70(S)**, 42.
27. Cassidy S, Eyres RD, Thomas NB, Whitehouse R, Williams PL (1999). A comparative study of the reporting accuracy of clinical radiographers and radiology registrars. *British Journal of Radiology*, **72(S)**, 19.
28. Piper KJ, Paterson AM, Ryan CM (2000). The implementation of a Radiographic Reporting Service for trauma examinations: final analyses of accuracy. *British Journal of Radiology*, **73(S)**, 91.

29. Sonnex EP, Tasker AD, Coulden RA (2001). The role of preliminary interpretation of chest radiographs by radiographers in the management of acute medical problems within a cardiothoracic centre. *British Journal of Radiology*, **75(S)**, 34.
30. Carter S, Manning D (1999). Performance monitoring during postgraduate radiography training in reporting – a case study. *Radiography*, **5**, 71-78.
31. Pitchers M (2002). Radiographic interpretation by Emergency Nurse Practitioners and Radiographers in a Minor Injuries Unit. BSc Dissertation.
32. Wolfe G (2002). How well do radiographers identify and describe abnormalities on axial and appendicular radiographs in A&E X-ray. BSc Dissertation.
33. Hargreaves JS, MacKay S (2000). The accuracy of the red dot: will it improve training? *British Journal of Radiology*, **73(S)**, 32.
34. Snaith BA (2000). Has radiography outgrown the red dot? *British Journal of Radiology*, **73(S)**, 32.

References of studies excluded

1. Cheyne WN, Field-Boden Q, Wilson J, Hall R (1987). The radiographer and frontline diagnosis. *Radiography*, **53**, 114.
2. Loughran CF (1996). Reporting of accident radiographs by radiographers. *RAD Magazine*, **2**, 34.
3. Berman L, de Lacey G, Twomey E, Twomey B, Welch T, Eban R (1986). Reducing errors in the accident department: a simple method using radiographers. *Radiography*, **52**, 143-4.
4. Boynes S, Scally AJ, Webster AJ, Kay K (1997). Radiographer reporting of the axial and appendicular skeleton by radiographers and nurse practitioners. *British Journal of Radiology*, **70(S)**, 123.
5. Piper K (1997). Reporting by radiographers: Findings of an Accredited Postgraduate Skeletal Reporting Programme in the UK. *Radiology*, **205**, 360.
6. Piper KJ, Paterson AM, Ryan CM (1998). The implementation of a radiographer reporting service for trauma examinations of the skeletal system. *British Journal of Radiology*, **71(S)**, 126.
7. Renwick I, Butt WP, Steele B (1990). An assessment of triage of casualty radiographs by radiographers. *British Journal of Radiology*, **63(S)**, 7.
8. Loughran CF (1994). Reporting of accident and emergency radiographs by radiographers: a study to determine the effectiveness of a training programme.

British Journal of Radiology, **67(S)**, 93.

9. Carr D, Gale A, Mugglestone (1997). Visual search strategies of radiographers: Evidence for role extension. *European Radiology*, **7(S)**, 179.
10. Carr D, Gale A, Mugglestone MD (1996). Can radiographers interpret medical images effectively? *British Journal of Radiology*, **69(S)**, 283.
11. Gale AG, Vernon J, Millar K, Worthington BS (1990). Reporting in a flash. *British Journal of Radiology*, **63(S)**, 71.
12. Quagehebeur GM, Shute M (1992). Audit of casualty x-rays. *British Journal of Radiology*, **65(S)**, 12.
13. Henderson PI, Lovegrove MJ, Hughes JW (1996). Radiographer reporting of the appendicular skeleton: initial evaluation of a pilot programme. *British Journal of Radiology*, **69(S)**, 128.
14. Piper K (1996). Accuracy of radiographers' reports in examinations of the skeletal system. *British Journal of Radiology*, **69(S)**, 282.
15. Piper K, Paterson AM (1997). The accuracy of radiographer's reports in examinations of the skeletal system. *British Journal of Radiology*, **70(S)**, 123.
16. Tasker AD, Sonnex E, Coulden RA (2000). The role of preliminary interpretation of surgical chest radiographs by radiographers in a cardiothoracic centre. *British Journal of Radiology*, **73(S)**, 32.
17. Manning D, Leach J, Bunting S (2000). A comparison of expert and novice performance in the detection of simulated pulmonary nodules. *Radiography*, **6**, 111-116.
18. Robinson PJA, Culpan G, Wiggins M (1999). Interpretation of selected accident and emergency radiographer examinations by radiographers: a review of 11000 cases. *British Journal of Radiology*, **72**, 546-51.
19. Tam CL, Chandramohan M, Thomas N (2002). A 2-year prospective study of radiographers for unsupervised reporting of casualty trauma radiographs. *British Journal of Radiology*, **75(S)**, 34.

Annex 2.2 Data extraction form

PLEASE TICK THE APPROPRIATE BOX(ES)

Reference Number:
Checklist Version: Final
Details of study:
Author(s)
Title

Source (e.g. Journal, Conference) Year / Volume / Pages

Institutional Affiliation (first author) and/or contact address

Publication bias (distribution of positive and negative findings):
Was the study published?
- ☐ Yes (e.g. paper)
- ☐ No (grey literature)

PART 1: Study eligibility and design

A Study eligibility:

A1 *Inclusion criteria*

For a study to be eligible for inclusion it must satisfy the criteria below:
- ☐ Radiographer(s) were compared with a reference standard to assess their plain radiograph reading performance
- ☐ Must include or have the potential to calculate an appropriate statistic that reflects accuracy (e.g. sensitivity, specificity).

A2 *Exclusion criteria*

A study will be excluded if:
- ☐ Images from other modalities (e.g. mammograms, ultrasound scans)
- ☐ Not performed during 1971-2002/10
- ☐ Case reports
- ☐ Visual search strategy study
- ☐ Duplication of data

A3 Is the study eligible (please explain why below)?
- ☐ Yes
- ☐ No

B Study design:

B1 In what setting was the study conducted?

- ☐ **outside of routine clinical practice** e.g. postgraduate course (which will be a study of the efficacy of the film reading performance of radiographers).
- ☐ **during routine clinical practice** (which will be a study of the effectiveness of the film reading performance of radiographers).

B2 What was the design of the study as an assessment of the film reading performance of a cohort(s) of observers?

- ☐ **Cohort A *versus* reference standard:** How accurate is cohort A when interpreting plain films?
- ☐ **Cohort A *versus* Cohort B *versus* reference standard:** How accurate is cohort A when interpreting plain films? How accurate is cohort B when interpreting plain films? How does cohort A compare to cohort B when interpreting plain films?
- ☐ **Cohort A *versus* Cohort B *versus* Cohort C *versus* reference standard:** How accurate is cohort A when interpreting plain films? How accurate is cohort B when interpreting plain films? How accurate is cohort C when interpreting plain films? Is there any difference in performance between the cohorts studied?

B3 What was the design of the study as described below?

- ☐ *diagnostic accuracy*: to assess the film reading performance of one (or more) group of observers in controlled (ideal) conditions.
- ☐ *diagnostic performance*: to assess the film reading performance of one group of observers during clinical practice.
- ☐ *diagnostic outcome*: to assess the film reading performance of two (or more) group of observers during clinical practice.

B4 What was the focus of the study with regards to the role of the observers being evaluated?

- ☐ *Pattern recognition study:* recognition of the presence of an abnormality (e.g. red dot system); or
- ☐ *Reporting study*: ability to produce a precise diagnosis (e.g. correct abnormality and location) using a combination of codes and free text.
- ☐ other (specify below).

PART 2: Quality criteria checklist

The quality criteria checklist has been subdivided into two sections: identification of biases and general methodological factors.

Section 1: Identification of biases

Each criterion is scored as:

- ☐ DONE (A) - there is evidence from an (un)published report or via personal communication that the criterion was achieved.
- ☐ NOT CLEAR (B) - if there is insufficient information from an (un)published report or via personal communication that the criterion was achieved. Missing information will be sought by the main reviewer.
- ☐ NOT DONE (C) - there is evidence from an (un)published report that the criterion was not achieved; or there is evidence from personal communication that the criterion was not achieved.
- ☐ Not applicable (N/A) - the criterion that the question is addressing is clearly not relevant to the particular study.

Can you please record the score you chose for each criterion by ticking the relevant box. Please record why you chose that score for each criterion under *Comment*:

Subjects [external validity]

If the study was conducted outside of routine clinical practice then answer section **A**. If the study was conducted during routine clinical practice then answer section **B**. Answer Section **C** for all studies to judge whether observers were appropriately selected.

Film selection

A Studies conducted outside of clinical practice (film cohort bias: spectrum; film filtering: eligibility criteria)

A1 Is spectrum bias present?

- ☐ Score DONE (A) if an attempt was made to include a non-random case mix based on at least three of the following factors: prevalence of disease, severity of disease, range of disease type, pertinent areas of the body; or
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if:
  - there was no record of the case mix of the films
  - two or less factors were taken into consideration when generating the case mix.
- ☐ N/A.

Comment:

A2 Are specific eligibility criteria stated for those included / excluded (*film filtering bias*)?

- ☐ Score DONE (A) if criteria are reported for all those films that were eligible for inclusion or exclusion from the study and the total number of films included is given as well as the number included/ excluded.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if criteria or numbers are not reported.
- ☐ N/A.

Comment:

B Studies conducted during clinical practice (referral biases: centripetal, popularity; film cohort: population; film filtering: eligibility criteria, film selection)

Questions B1-2 provide only information. A judgement from this information is required to assess the presence or absence of these referral biases.

B1 Is the establishment(s) where the study was undertaken stated (*centripetal bias*)?

- ☐ Score DONE (A) if the establishment is the place of origin of the study.
- ☐ Score NOT DONE (C) if not reported.

Comment:

B2 Is the establishment from where the patients were referred stated (*popularity bias*)?

- ☐ Score DONE (A) if the establishment is clearly stated e.g. A&E department.
- ☐ Score NOT DONE (C) if not reported.

Comment:

B3 Is population bias present?

- ☐ Score DONE (A) if:
- a series of films over a suitable time period was included; <u>or</u>
- a valid random sample of films were selected in a way so that the professionals responsible for interpreting the films had no choice as to what films they interpreted and the random process is described explicitly, e.g. the use of random number tables.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if:
- there is no statement as to the length of the time period during which the consecutive series of films were interpreted; or the series of films that were included was not during a long enough time period.
- the allocation procedure for randomisation is not described; or alternation such as reference to case record numbers, dates of birth, day of the week or any other such approach was used in the selection of films.
- ☐ N/A.

Comment:

B4 Are specific eligibility criteria stated for those included / excluded?

- ☐ Score DONE (A) if criteria are reported for all those films that were eligible for inclusion or exclusion from the study and the total number of films included is given as well as the number included/ excluded; or is it clear that all films were included.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if criteria or numbers are not reported.
- ☐ N/A.

Comment:

B5 Is film selection bias present?

- ☐ Score DONE (A) if:
- all films eligible to be included in the study were interpreted by the observers under evaluation; and
- observers could not choose which eligible films to interpret.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if:
- not all the eligible films were interpreted by the observers; or
- the observers could choose which eligible films to interpret (i.e. systematic exclusions).
- ☐ N/A.

Comment:

Observer selection

C Relevant to all studies

C1 Is observer cohort bias present?

- ☐ Score DONE (A) if an appropriate group of observers were selected.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if an inappropriate group of observers were selected.
- ☐ N/A.

Comment:

C2 Is observer cohort comparator bias present?

- ☐ Score DONE (A) if the study group (received training) and control group (no training) were matched according to the following characteristics: professional group; number of years experience in the profession; number of years experience in a relevant speciality (e.g. A&E); number of years experience interpreting images (e.g. ultrasound).
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if the study group and control group were not matched according to the above characteristics.
- ☐ N/A.

Comment:

Study [internal validity]

All studies should be assessed in relation to the following criteria:

D Application of the reference standard

D1 Is verification bias present?

- Score DONE (A) if all the films interpreted by the observers under evaluation were also interpreted by the reference standard or a correction is performed by the authors.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if not all films interpreted by the observers under evaluation were also interpreted by the reference standard.
- N/A.

Comment:

D2 Is work-up bias present?

- Score DONE (A) if the interpretation made by the observers under evaluation is not used to decide whether the reference standard is applied or a correction is performed by the authors.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if the interpretation made by the observers under evaluation is used to decide whether the reference standard is applied.
- N/A.

Comment:

D3 Is incorporation bias present?

- Score DONE (A) if the interpretation of an observer under evaluation is not incorporated into the evidence used to diagnose the disease or is itself not used as the reference standard.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if the interpretation of an observer under evaluation is incorporated into the evidence used to diagnose the disease or is itself used as the reference standard.
- N/A.

Comment:

E Measurement of results (disease progression; withdrawal bias: indeterminate observer interpretations, follow-up; observer variability: inter-observer; intra-observer; arbiter variability: inter-arbiter, intra-arbiter).

E1 Is disease progression bias present?
- ❑ Score DONE (A) if appropriate radiological and clinical review is used.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if inappropriate clinical and radiological review is used.
- ❑ N/A.

Comment:

E2 Are there any indeterminate (i.e. equivocal, non-diagnostic) observer interpretations?
- ❑ Score DONE (A) if all films and subsequent interpretations are included irrespective of their indeterminability.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if films are excluded due to indeterminate interpretations.
- ❑ N/A.

Comment:

E3 Are there any patients lost to follow-up?
- ❑ Score DONE (A) if all films and clinical information is available for verification.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if patients are excluded or films not reported owing to loss.
- ❑ N/A.

Comment:

E4 Is any attempt made to assess intra-observer variability?
- ❑ Score DONE (A) if for a subsample of the films interpreted data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

E5 Is any attempt made to assess inter-observer variability?
- ❑ Score DONE (A) if for a subsample of the films interpreted data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

E6 Is any attempt made to assess intra-arbiter variability?

- ❑ Score DONE (A) if for a subsample of the interpretations compared data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

E7 Is any attempt made to measure inter-arbiter variability?

- ❑ Score DONE (A) if for a subsample of the interpretations compared data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

F Independence of interpretations

F1 Is observer review bias present?

- ❑ Score DONE (A) if the observers being evaluated were blinded or unaware of the interpretation made by the reference standard when interpreting the films.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the observers being evaluated were aware of the interpretation made by the reference standard when interpreting the films.
- ❑ N/A.

Comment:

F2 Is reference standard review bias present?

- ❑ Score DONE (A) if the reference standard was blinded or unaware of the interpretation made by the observers under evaluation.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the reference standard was aware of the interpretation made by the observers under evaluation.
- ❑ N/A.

Comment:

F3 Is observer bias present?

- ❑ Score DONE (A) if all observers always interpreted the films independent of each other.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if observers did not always interpret the films independent of each other.
- ❑ N/A.

Comment:

F4 Is observer comparator bias present?

- Score DONE (A) if all observers interpreted the same or a similar set of films independent of each other.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if observers did not always interpret the same or a similar set of films independent of each other.
- N/A.

Comment:

F5 Is co-image bias present?

- Score DONE (A) if all observers only had access to the films that they were being asked to interpret and not images from other modalities in relation to the same examination.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if observers had access to images from other modalities in relation to the films that they were being asked to interpret.
- N/A.

Comment:

F6 Is arbiter review bias present?

- Score DONE (A) if the arbiter was not one of the observers under evaluation or the reference standard.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if the arbiter was one of the observers under evaluation and/or the reference standard.
- N/A.

Comment:

F7 Is arbiter bias present?

- Score DONE (A) if the arbiter was blind or unaware as to whether the report was made by an observer under evaluation or the reference standard.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if the arbiter was aware of who was responsible for either of the reports.
- N/A.

Comment:

F8 Is film access bias present?

- Score DONE (A) if the arbiter judged whether interpretations agreed or not without access to the films.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if the arbiter made use of the films during the process of judging whether interpretations agreed or not.
- N/A.

Comment:

ADDITIONAL VALIDITY CRITERIA FOR STUDIES COMPARING TWO (OR MORE) COHORTS

F9 Is cohort comparator bias present?
- ☐ Score DONE (A) if the cohorts of observers interpreted the same films independent of each other.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if the cohorts of observers did not always interpret the films independent of each other; or did not report on the same films.
- ☐ N/A.

Comment:

F10 Is co-image comparator bias present?
- ☐ Score DONE (A) if both cohort of observers had similar access to the relevant plain films and did not have access to images from other modalities.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if both cohort of observers did not have similar access to the relevant plain films and one cohort of observers had access to images from other modalities.
- ☐ N/A.

Comment:

F11 Is arbiter comparator bias present?
- ☐ Score DONE (A) if the arbiter was blind or unaware as to who was responsible for the interpretations when judging whether they agreed or not.
- ☐ Score NOT CLEAR (B) if there is insufficient information.
- ☐ Score NOT DONE (C) if the arbiter was aware of who was responsible for the interpretations when judging whether they agreed or not.
- ☐ N/A.

Comment:

Section 2: General methodological standards

Each criterion is scored as:

- ❑ DONE (A) - there is evidence from an (un)published report or via personal communication that the criterion was achieved.
- ❑ NOT DONE (C) - there is no evidence from an (un)published report that the criterion was achieved; or there is evidence from personal communication that the criterion was not achieved; or there is no evidence from an (un)published report or via personal communication that the criterion was achieved.
- ❑ Not applicable (N/A) - the criterion that the question is addressing is clearly not relevant to the particular study.

Can you please record the score you chose for each criterion by ticking the relevant box(es). Please record why you chose that score for each criterion under *Comment*:

G Subjects (films)

G1 Was an appropriate sample size considered?

- ❑ Score DONE (A) if the study:
- • measured the performance of a single cohort of observers and the sample size was calculated according to how precise an estimate of the sensitivity and specificity was required.
- • reports an attempt to calculate the sample size required to detect clinically important effects as statistically significant between two (or more) cohorts of observers, and if possible, record the power under *Comment*.
- ❑ Score NOT DONE (C) if:
- • no reference is made to the sample size required.
- • no power calculation is stated, or the study did not attempt to calculate the sample size required.
- ❑ N/A.

Comment:

H Study

H1 Was a normal/abnormal report adequately defined?

- ❑ Score DONE (A) if an explicit attempt was made to adequately define a normal/abnormal report.
- ❑ Score NOT DONE (C) if a normal/abnormal report was not adequately defined.
- ❑ N/A.

Comment:

H2 Was the performance of the observers placed in the context of the diagnostic sequence (i.e. referral filters e.g. red dot system, casualty officers [cold], hot)?

- ❑ Score DONE (A) if the study made an explicit attempt to report the process through which the films had passed before they were interpreted by the observers under evaluation.
- ❑ Score NOT DONE (C) if the study did not report the context in which the films were interpreted.
- ❑ N/A.

Comment:

H3 If the combined performance of two (or more) different groups of observers is assessed was the contribution of the individual groups to the overall validity of the combination of groups determined?

- ☐ Score DONE (A) if every single group within a combination of groups was evaluated.
- ☐ Score NOT DONE (C) if not every single group of a combination of groups were evaluated.
- ☐ N/A.

Comment:

H4 Was an appropriate (valid) reference ("gold" or "criterion") standard used?
Score DONE (A) if the study reported a suitable reference standard:

- ☐ A1: a double/triple blind consultant radiological report.
- ☐ A2: a single consultant radiological report that was validated in an acceptable way e.g. via clinical follow-up.
- ☐ A3: a single consultant radiological report that was not validated.
- ☐ Score NOT DONE (C) if an inappropriate reference standard is reported e.g. a combination of radiologists at different grades, the observers under evaluation were also used as the reference standard or included in the process of generating the reference standard; not reported in the paper.
- ☐ N/A.

Comment:

H5 Was an appropriate (valid) arbiter used?
Score DONE (A) if the study used a suitable arbiter:

- ☐ A1: external: panel
- ☐ A2: external: consultant radiologist
- ☐ A3: internal: panel
- ☐ A4: internal: consultant radiologist
- ☐ A5: radiographer(s) trained to report and if unsure an independent consultant radiologist.
- ☐ A6: untrained radiographer(s) and if unsure an independent consultant radiologist.
- ☐ Score NOT DONE (C) if: the study reported an inappropriate arbiter e.g. independent untrained radiographer(s) with no referral to radiologist, a person under evaluation is responsible for comparing the reports; not reported in the paper.
- ☐ N/A.

Comment:

H6 Was a control used in the study (*appropriate choice of control activity*)?

- ☐ Score DONE (A) if an appropriate control was used within the context of the particular study.
- ☐ Score NOT DONE (C) if an inappropriate control was used; or a control was appropriate but not used.
- ☐ N/A.

Comment:

I Interpretation

I1 Were films appropriately analysed for pertinent subgroups?

- ❑ Score DONE (A) if an attempt was made to analyse the observers performance for pertinent medical subgroups, e.g. areas of the body.
- ❑ Score NOT DONE (C) if there was no attempt to analyse pertinent medical subgroups.
- ❑ N/A.

Comment:

I2 Was the data presented in enough detail to allow for the calculation of appropriate indices of performance (e.g. sensitivity and specificity) and confidence intervals?

- ❑ Score DONE (A) if the data was presented in enough detail to calculate the above.
- ❑ Score NOT DONE (C) if the data was NOT presented in enough detail to calculate the above.
- ❑ N/A.

Comment:

I3 Are indeterminate observer interpretations appropriately presented?

- ❑ Score DONE (A) if a study reported:
- • all of the appropriate positive, negative and indeterminate interpretations; and
- • whether indeterminate interpretations had been included or excluded when indices of performance were calculated.
- ❑ Score NOT DONE (C) if the study did not:
- • attempt to categorise reports as positive, negative, and indeterminate.
- • state whether indeterminate results had been included or excluded when indices of performance were calculated.
- ❑ N/A.

Comment:

PART 3: Factor checklist

The purpose of this section is to extract relevant factual information regarding the professionals, patients (films) and outcomes included in the study, as well as other relevant issues.

A Characteristics of professionals

Please tick next to which cohort(s) of professionals were under evaluation and record the number of observers within each cohort in the brackets [ ].

A1 Profession:	Number of observers
❑ Radiographer (not trained to report)	[ ]
❑ Radiographer (during training)	[ ]
❑ Radiographer (pre-training)	[ ]
❑ Radiographer (post-training)	[ ]
❑ Radiologist	[ ]
❑ Casualty Officer	[ ]
❑ Nurse practitioner	[ ]
❑ Other (specify)	[ ]
❑ NOT CLEAR	

A2 Personal skills: Please record below the time since graduation (or years in practice) as well as seniority (score NOT CLEAR if no information is available) of all the professionals involved in the study.

A3 Format:

(i) What was the method used by the observers to interpret the films e.g. abnormal, normal or equivocal; combination of codes and free text (and the definitions)? State NOT CLEAR if no explanation of the method used is given.

(ii) What was the method used by the reference standard to interpret the films? State NOT CLEAR if no explanation of the method used is given.

B Characteristics of the patients

B1 Please tick next to which area(s) of the body were included:

- ☐ upper limb (shoulder girdle to fingers)
- ☐ lower limb (pelvis to toes)
- ☐ chest (excluding thoracic cage)
- ☐ thoracic cage
- ☐ facial bones
- ☐ skull
- ☐ abdomen
- ☐ spine
- ☐ foreign bodies
- ☐ NOT CLEAR
- ☐ Other (specify below)

B2 Please tick next to which type of plain films were included and record the number of films in the brackets [ ]:

	Number of films
☐ accident and emergency	[ ]
☐ general practitioner	[ ]
☐ NOT CLEAR	[ ]
☐ Other (specify below)	[ ]

B3 Please tick next to which setting(s) the films were interpreted in and record the number of individual settings in the brackets [ ]:

	Number
☐ University based/teaching hospital	[ ]
☐ Non-teaching hospital	[ ]
☐ Mixed hospital	[ ]
☐ District General Hospital	[ ]
☐ Infirmary	[ ]
☐ Minor Injuries Unit	[ ]
☐ Community based	[ ]
☐ Postgraduate course (exam conditions)	[ ]
☐ NOT CLEAR	[ ]
☐ Other (specify below)	[ ]

C Outcomes

State the results of the performance of each observer or cohort of observers in an appropriate form (e.g. sensitivity, specificity, diagnostic accuracy and prevalence). Include statistical significance if reported for these differences, only if the units of analysis are the same. State the p value, significant tests used and appropriateness, as well as measures of variability e.g. confidence intervals. Insert the raw data when possible into a 2x2 contingency table or other suitable table. Having collected all relevant results, they will either be entered in the review as they are, or transformed into something more useful.

D Other relevant issues

PLEASE TICK THE APPROPRIATE BOX FOR EACH QUESTION

D1 Was an attempt made to demonstrate how the interpretation of the observer(s) reduced uncertainty in diagnosis for the referring clinician?

- ☐ Yes
- ☐ No
- ☐ NOT CLEAR

D2 Was an explicit attempt made to assess the effect of discrepancies between interpretations of the observer(s) in comparison with the reference standard on patient management and/or outcome?

- ☐ Yes
- ☐ No
- ☐ NOT CLEAR

D3 Was an attempt made to conduct an economic analysis?

- ☐ Yes
- ☐ No
- ☐ NOT CLEAR

Explanatory notes for the checklist

Unless otherwise specified, the word 'films' in the data collection checklist is synonymous with plain film x-ray examinations and the word 'observer(s)' refers to the individual professional(s) for which performance is being assessed. The word 'cohort' refers to a group of observers from one profession (e.g. radiographers, radiologists, nurse practitioners), or to a specific process which may include one or more professional groups that simulates a cohort (e.g. radiographers and nurse practitioners combined). The word 'arbiter' refers to the person or people responsible for comparing the interpretations for concordance. The following list further clarifies some of the questions asked in the checklist.

Part 2: Quality criteria checklist

Section 1: Identification of biases in the overall design of the study

Subjects [internal and external validity]

Film selection

A Studies conducted outside of clinical practice

A1 *Spectrum bias* - this is present when not all of the following factors are considered when selecting the sample of films: prevalence of disease, severity of disease, disease type, and areas of the body.

A2 *Film filtering bias* - this is present if there is no record of the criteria used to determine which films were eligible for inclusion or exclusion. This bias is also present if the total number of films is not given and the number included/ excluded.

B Studies conducted during clinical practice

B1 *Centripetal bias* - this is present if there is no record of the establishment where the study was undertaken.

B2 *Popularity bias* - this is present if the establishment from where patients were referred is not clearly stated.

B3 *Population bias* - this is present if a series of films included in a sample was not over a suitable time period or was not a valid random sample. The decision as to whether the observers interpreted a series of films over a long enough time period is a subjective one.

B4 *Film filtering bias* - see A2.

B5 *Film selection bias* - this occurs if the observers under evaluation do not interpret all the films that are eligible to included in the study and/ or have the opportunity to choose which eligible films they want to interpret.

Observer selection

C Relevant to all studies

C1 *Observer cohort bias* - this occurs if an inappropriate selection of observers are included in a study with regards to the research question that is being addressed.

C2 *Observer cohort comparator bias* - this occurs if two (or more) groups of observers are compared without the appropriate use of matching. For studies that assess the effectiveness of a training programme and are comparing a study group (receive training) with a control group (no training) the two groups should be matched for the characteristics listed to ensure comparability.

Study [internal validity]

D Application of the reference standard

D1 *Verification bias* - this occurs when not all of the films interpreted by the observers under evaluation are interpreted by the same reference standard for any reason e.g. economic limitations, decisions based on clinical signs and symptoms.

D2 *Work-up bias* - this occurs when not all the films receive definitive confirmation with the reference standard due to the interpretation of the observers under evaluation. Using this definition, if work-up bias is present then verification bias is also present but not *vice versa.*

D3 *Incorporation bias* - this occurs if the report of an observer under evaluation is incorporated into the evidence used to diagnose the disease. This also occurs if the report of the observer under evaluation is used as the reference standard e.g. the report of an observer under evaluation within a cohort, such as a radiologist, is used as the reference standard. Incorporation bias is not present if the study is designed to follow the progression of a disease, and a definitive endpoint reference standard is used for diagnosis.

E Measurement of results

E1 *Disease progression bias* - this occurs if there is a long time period between the initial report and subsequent clinical follow up. If the reference standard only involves reporting films then this bias is not applicable. However, if the reference standard includes clinical follow-up, it is important that there is appropriate radiological review. This is to ensure that the initial film, for example, was incorrectly interpreted by an observer because of a missed overt fracture rather than the film being correctly reported but an occult fracture resulted in the patient re-attending.

E2 *Indeterminate interpretation bias* - this is present if not all indeterminate interpretations are included when measuring observers performance. If films are excluded for this reason prior to the application of the reference standard this will introduce work-up bias.

E3 *Loss to follow-up bias* - this occurs if information is systematically lost so that the reference standard can not be applied.

E4 *Intra-observer variability bias* - this occurs if the observers under evaluation did not re-interpret a subsample of the films to measure their consistency in the interpretation of films.

E5 *Inter-observer variability bias* - this occurs if the observers within a cohort did not report on the same subsample of films. If only one observer this is not applicable.

E6 *Intra-arbiter variability bias* - this occurs if the same arbiter did not re-apply the criteria used to judge whether there is concordance between interpretations on a subsample of cases.

E7 *Inter-arbiter variability bias* - this occurs if two independent arbiters did not compare a subsample of the observer interpretations with the reference standard to assess whether the criteria was applied consistently by different people.

F Independence of interpretations

F1 *Observer review bias* - this occurs if the observers being evaluated are aware of the interpretation made by the reference standard when interpreting the films. If the reference standard used is clinical follow-up, so long as it is not a retrospective study the results of the definitive diagnosis must be unknown at the time of the interpretation by the observers under evaluation. Thus, the bias is absent.

F2 *Reference standard review bias* - this occurs if the interpretations of the observers under evaluation are known when the diagnosis is made by the reference standard.

F3 *Observer bias* - this occurs if the individual observers within a cohort do not interpret the films independent of each other.

F4 *Observer comparator bias* - this occurs if an attempt is made to compare the performance of observers within a cohort and not all observers interpreted the same or a similar set of films independent of each other.

F5 *Co-image bias* - this occurs if additional images were available to a cohort of observers other than those they were being assessed to interpret with the exception of previous plain films.

F6 *Arbiter review bias* - this occurs if the arbiter was one of the observers under evaluation or was the reference standard.

F7 *Arbiter bias* - this occurs if the arbiter was aware as to whether the interpretation was made by the observer(s) under evaluation or the reference standard.

F8 *Film access bias* - this occurs if the arbiter had access to films whilst judging whether interpretations agreed or not. Their interpretation can incorrectly influence the decision as to whether the reports agree or not, or as to which report is correct.

ADDITIONAL VALIDITY CRITERIA FOR STUDIES COMPARING TWO (OR MORE) COHORTS

F9 *Cohort comparator bias* - this occurs if the cohorts of observers did not interpret the same films independent of each other. For example, a study may have compared radiographers performance with the reference standard and radiologists performance with the reference standard. Both the cohort of radiographers and radiologists should interpret the films independently. Furthermore, the two cohorts should report on the same or a comparable batch of films.

F10 *Co-image comparator bias* - this occurs if one cohort of observers had access to images from other modalities.

F11 *Arbiter comparator bias* - this occurs if the arbiter was aware as to which of the interpretations was made by the different cohort of observers.

Section 2: General methodological factors

G Subjects (films)

G1 (a) If the study is measuring the performance of a single cohort of observers the sample size should be calculated according to how precise an estimate of the sensitivity and specificity is required. (b) Studies comparing cohorts should make use of a power calculation.

H Study

H1 Whether the definition of normal or abnormal is acceptable is a subjective one. The important issue is whether a definition was available.

H2 It is important that a study describes the diagnostic sequence through which films pass as this will affect the case mix of films that the observers interpret and subsequently the generalisability of the results. This criterion will not be applicable to postgraduate studies.

H3 Some studies may assess the combined performance of two groups of observers such as the interpretation made by a nurse practitioner having seen the interpretation made by a radiographer. This type of study should also assess the performance of the two groups separately to identify the contribution of each group to the combined effort.

H6 The relevant control may vary, if one is necessary, depending on the research question.

Annex 2.3 Checklist results from primary studies

Table A2.3.1 Results from quality criteria checklist for diagnostic accuracy studies

	Henderson 1999[14]	Wilson 1999[22]	Piper & Paterson 1997[17]	Callaway et al 1997[3]	Piper et al 2002[18]
A1	A	A	A	A	A
A2	C	A	A	A	A
C1	A	A	A	A	A
C2	N/A	A	A	N/A	N/A
D1	A	A	A	A	A
D2	A	A	A	A	N/A
D3	A	A	A	A	A
E1	N/A	N/A	N/A	N/A	N/A
E2	A	A	A	A	A
E3	N/A	N/A	N/A	N/A	N/A
E4	C	C	C	C	N/A
E5	C	C	C	C	A
E6	C	C	C	C	C
E7	C	C	A	C	N/A
F1	A	A	A	A	A
F2	A	A	A	A	A
F3	A	A	A	A	A
F4	A	N/A	N/A	N/A	A
F5	A	A	A	A	A
F6	A	A	A	A	A
F7	B	B	C	C	C
F8	C	C	C	C	A
F9	N/A	A	A	A	N/A
F10	N/A	A	A	A	N/A
F11	N/A	A	C	C	N/A
G1	C	C	C	C	C
H1	A	A	A	A	A
H2	N/A	N/A	N/A	N/A	N/A
H3	N/A	N/A	N/A	N/A	N/A
H4	A3	A1	A1	A3	A1
H5	A6	A6	A3	A3	A2
H6	N/A	A	N/A	N/A	N/A
I1	C	A	A	N/A	A
I2	C	C	A	N/A	A
I3	C	N/A	N/A	A	N/A

Table A2.3.1 Results from quality criteria checklist for diagnostic accuracy studies

	Boynes *et al* 1999[10]	Seymour & White 1996[23]	McConnell & Webster 1999[13]	McMillan *et al* 1995[11]	Hughes *et al* 1996[2]	Cassidy *et al* 1995[27]
A1	A	A	A	A	A	A
A2	A	A	A	A	B	B
C1	A	C	A	A	A	A
C2	A	N/A	N/A	N/A	N/A	N/A
D1	A	A	A	A	A	A
D2	A	A	A	A	A	A
D3	A	A	A	A	A	A
E1	N/A	N/A	N/A	N/A	N/A	N/A
E2	A	A	A	A	A	B
E3	N/A	N/A	N/A	N/A	N/A	N/A
E4	C	C	N/A	C	C	C
E5	C	C	C	C	C	C
E6	C	C	C	C	C	C
E7	C	C	C	C	C	C
F1	A	A	A	A	A	A
F2	A	A	A	A	A	A
F3	A	A	A	C	B	A
F4	N/A	A	N/A	N/A	N/A	N/A
F5	A	A	A	A	B	A
F6	A	C	A	A	B	A
F7	C	C	C	C	B	C
F8	C	C	C	A	B	C
F9	A	A	N/A	N/A	N/A	A
F10	A	A	N/A	N/A	N/A	A
F11	C	C	N/A	N/A	N/A	C
G1	C	C	C	C	C	C
H1	A	A	A	A	A	A
H2	N/A	N/A	N/A	N/A	N/A	N/A
H3	N/A	N/A	N/A	N/A	N/A	N/A
H4	A1	A3	A3	A1	A3	A1
H5	A5	A4	A5	C	C	A6
H6	A	N/A	C	N/A	C	N/A
I1	C	N/A	C	A	C	N/A
I2	A	C	C	C	A	N/A
I3	N/A	N/A	N/A	A	C	N/A

Table A2.3.2 Results from quality criteria checklist for diagnostic performance studies

	Loughran *et al* 1996[19]	Raynor 1999[15]	Manning 1999[12]	Eyres & Williams 1999[9]	Piper *et al* 2000[28]
B1	A	A	A	A	A
B2	A	A	A	A	A
B3	A	A	A	A	A
B4	A	A	A	A	A
B5	A	A	A	A	A
C1	A	A	A	A	A
C2	N/A	N/A	N/A	N/A	N/A
D1	A	A	A	A	A
D2	A	A	A	A	A
D3	A	A	A	C	C
E1	N/A	N/A	N/A	N/A	N/A
E2	N/A	N/A	A	A	C
E3	A	A	A	B	A
E4	C	C	C	C	C
E5	C	N/A	C	C	C
E6	C	C	N/A	C	C
E7	C	C	N/A	C	N/A
F1	A	A	C	A	A
F2	C	C	A	A	C
F3	A	A	B	A	A
F4	N/A	N/A	N/A	N/A	N/A
F5	A	A	A	A	A
F6	C	C	C	C	C
F7	C	C	C	C	C
F8	C	C	A	B	C
G1	C	C	C	C	A
H1	A	A	A	A	A
H2	A	A	A	C	A
H3	N/A	N/A	N/A	N/A	N/A
H4	A3	A3	A3	C	C
H5	C	C	C	C	A3
H6	N/A	N/A	N/A	N/A	N/A
I1	C	C	A	A	A
I2	C	C	A	A	A
I3	N/A	N/A	N/A	N/A	A

Table A2.3.2 Results from quality criteria checklist for diagnostic performance studies

	Webster & Gallacher 1998[26a]	Ford & Crawshaw 1999[21]	Renwick *et al* 1991[7]	Eyres & Williams 1997[8]	Bowman 1991[25]
B1	A	A	A	C	A
B2	A	A	A	A	A
B3	A	C	A	A	C
B4	A	C	A	A	C
B5	A	A	C	A	A
C1	A	A	A	A	A
C2	N/A	N/A	N/A	N/A	N/A
D1	A	C	A	A	A
D2	A	C	A	A	A
D3	A	A	A	C	A
E1	N/A	N/A	N/A	N/A	N/A
E2	A	C	A	A	B
E3	A	A	A	B	B
E4	C	C	C	C	C
E5	N/A	C	C	C	C
E6	C	C	C	C	C
E7	C	C	C	C	C
F1	A	A	A	A	A
F2	C	A	C	A	C
F3	A	A	N/A	B	B
F4	N/A	N/A	N/A	A	B
F5	A	A	A	A	A
F6	C	C	B	B	B
F7	C	C	C	C	B
F8	C	A	B	C	B
G1	C	C	C	C	C
H1	A	A	A	A	C
H2	A	A	A	A	A
H3	N/A	N/A	N/A	N/A	N/A
H4	A3	A2	C	C	A3
H5	A4	C	C	C	C
H6	C	N/A	N/A	C	C
I1	C	C	A	A	C
I2	C	A	A	A	A
I3	A	N/A	A	N/A	C

Table A2.3.2 Results from quality criteria checklist for diagnostic performance studies

	Wolfe 2002[32]	Sonnex *et al* 2001[29]	Hargreaves & MacKay 2000[33]	Carter & Manning 1999[30]	Snaith 2000[34]
B1	A	A	A	A	A
B2	A	A	A	A	A
B3	C	A	A	A	C
B4	A	A	B	A	A
B5	C	A	B	A	C
C1	A	A	B	A	C
C2	N/A	N/A	N/A	N/A	N/A
D1	A	A	B	A	C
D2	N/A	N/A	N/A	N/A	C
D3	C	A	B	A	C
E1	N/A	N/A	N/A	N/A	N/A
E2	A	A	N/A	A	A
E3	A	A	B	A	C
E4	N/A	N/A	N/A	N/A	N/A
E5	C	N/A	N/A	N/A	N/A
E6	C	C	C	C	C
E7	C	C	C	C	C
F1	A	A	A	A	A
F2	C	C	C	A	A
F3	C	N/A	N/A	N/A	N/A
F4	C	N/A	N/A	N/A	N/A
F5	A	B	A	N/A	A
F6	A	C	B	C	C
F7	C	C	C	C	A
F8	A	C	C	C	C
G1	A	C	C	C	C
H1	A	A	C	C	A
H2	A	A	A	A	A
H3	N/A	N/A	N/A	N/A	N/A
H4	C	C	C	A3	A3
H5	C	C	C	A4	A5
H6	N/A	N/A	C	C	N/A
I1	A	N/A	C	A	C
I2	A	A	C	A	A
I3	N/A	N/A	N/A	N/A	A

Table A2.3.3 Results from quality criteria checklist for diagnostic outcome studies

	Remedios *et al* 1998[1]	Berman *et al* 1985[6]	Balcam & Hood 1998[4]	Webster & Gallacher 1998[26]	Timmis & Burnett 1995[20]
B1	A	A	A	A	A
B2	A	A	A	A	A
B3	C	A	A	A	A
B4	C	A	A	A	A
B5	C	C	A	A	C
C1	A	A	A	A	A
C2	N/A	N/A	N/A	N/A	N/A
D1	A	A	C	A	A
D2	A	A	C	A	A
D3	A	A	A	A	A
E1	N/A	N/A	N/A	N/A	N/A
E2	C	C	A	A	A
E3	B	A	A	A	A
E4	C	C	C	C	C
E5	C	C	C	C	C
E6	C	C	C	C	C
E7	C	C	C	C	C
F1	A	A	A	A	A
F2	C	C	C	A	C
F3	N/A	B	B	A	N/A
F4	N/A	N/A	N/A	N/A	N/A
F5	A	A	A	A	A
F6	B	C	C	A	C
F7	B	C	C	B	C
F8	C	B	C	B	C
F9	N/A	B	A	A	C
F10	A	A	A	A	A
F11	C	C	B	B	C
G1	C	C	C	C	C
H1	A	A	A	A	A
H2	A	A	A	A	A
H3	C	N/A	N/A	N/A	N/A
H4	C	A3	A3	A3	A3
H5	A4	A3	A4	A4	A4
H6	N/A	N/A	N/A	C	N/A
I1	C	C	A	C	A
I2	A	A	A	C	C
I3	C	A	A	A	A

Table A2.3.3 Results from quality criteria checklist for diagnostic outcome studies

	Robinson 1996[5]	Giles 1989[24]	Loughan 1994[16]	Pitchers 2002[31]
B1	A	A	A	A
B2	A	A	A	A
B3	C	C	A	C
B4	A	A	A	A
B5	A	C	A	C
C1	A	A	A	A
C2	N/A	N/A	N/A	N/A
D1	C	A	C	A
D2	C	A	C	N/A
D3	A	A	B	A
E1	N/A	N/A	N/A	N/A
E2	A	A	A	A
E3	A	N/A	C	C
E4	B	C	C	N/A
E5	C	C	A	N/A
E6	C	C	C	C
E7	C	C	C	C
F1	A	A	A	A
F2	A	C	B	A
F3	N/A	N/A	B	N/A
F4	N/A	N/A	A	A
F5	A	A	A	A
F6	A	C	B	A
F7	C	C	C	C
F8	C	C	C	A
F9	N/A	N/A	A	A
F10	A	A	A	A
F11	C	C	C	C
G1	C	C	C	C
H1	A	N/A	A	A
H2	A	A	C	A
H3	N/A	A	N/A	N/A
H4	A2	A3	A3	A3
H5	A6	C	C	A6
H6	N/A	N/A	C	N/A
I1	A	C	C	N/A
I2	A	C	A	A
I3	A	C	A	N/A

Annex 2.4 Results of scoring systems

Table A2.4.1 Summary of the scoring systems

Criterion	SS1	SS2	SS3	SS4
A1	High = 3	High = 3	High = 3	High = 3
A2	High = 3	High = 3	High = 3	High = 3
B1	Low = 1	Low = 1	Low = 1	Low = 1
B2	Low = 1	Medium = 2	Low = 1	Medium = 2
B3	High = 3	Medium = 2	High = 3	Medium = 2
B4	High = 3	Medium = 2	High = 3	Medium = 2
B5	High = 3	High = 3	High = 3	High = 3
C1	Medium = 2	Medium = 2	High = 3	Medium = 2
C2	High = 3	High = 3	High = 3	High = 3
D1	High = 3	High = 3	High = 3	High = 3
D2	High = 3	High = 3	High = 3	High = 3
D3	High = 3	High = 3	High = 3	High = 3
E1	Medium = 2	Medium = 2	High = 3	Low = 1
E2	Medium = 2	Low = 1	Medium = 2	Low = 1
E3	Low = 1	Medium = 2	Medium = 2	Low = 1
E4	Low = 1	Low = 1	Low = 1	Low = 1
E5	Low = 1	Low = 1	Low = 1	Low = 1
E6	Medium = 2	Low = 1	Medium = 2	Low = 1
E7	Medium = 2	Medium = 2	Medium = 2	Medium = 2
F1	High = 3	High = 3	High = 3	High = 3
F2	Medium = 2	Medium = 2	High = 3	Low = 1
F3	High = 3	Medium = 2	High = 3	Medium = 2
F4	High = 3	Medium = 2	High = 3	Medium = 2
F5	Low = 1	Low = 1	Low = 1	Low = 1
F6	Medium = 2	Medium = 2	High = 3	Low = 1
F7	Medium = 2	Low = 1	Medium = 2	Low = 1
F8	Medium = 2	Medium = 2	High = 3	Low = 1
F9	High = 3	High = 3	High = 3	High = 3
F10	High = 3	Medium = 2	Medium = 2	High = 3
F11	Medium = 2	Medium = 2	Medium = 2	Medium = 2
G1	Medium = 2	Medium = 2	High = 3	Low = 1
H1	Medium = 2	Low = 1	Medium = 2	Low = 1
H2	Medium = 2	Low = 1	Medium = 2	Low = 1
H3	Medium = 2	Low = 1	Medium = 2	Low = 1
H4	• A1 = 0 • A2 = 1 • A3 = 2 • Not Done=3	• A1 = 0 • A2 = 1 • A3 = 2 • Not Done=3	• A1 = 0 • A2 = 1 • A3 = 2 • Not Done=3	• A1 = 0 • A2 = 1 • A3 = 2 • Not Done=3
H5	• A1 & 2 = 0 • A3 &4 =1 • A5 = 2 • A6 & Not Done = 3	• A1 & 2 = 0 • A3 &4 =1 • A5 = 2 • A6 & Not Done = 3	• A1 & 2 = 0 • A3 &4 =1 • A5 = 2 A6 & Not Done = 3	• A1 & 2 = 0 • A3 &4 =1 • A5 = 2 • A6 & Not Done = 3
H6	High = 3	High = 3	High = 3	High = 3
I1	Medium = 2	Low = 1	Medium = 2	Medium = 2
I2	Medium = 2	Low = 1	Medium = 2	Medium = 2
I3	Medium = 2	Low = 1	Medium = 2	Medium = 2
Maximum Total	91	79	97	77

Table A2.4.2 Ranking of studies for the different scoring systems

Rank	SS1		SS2		SS3		SS4	
	Study	Score	Study	Score	Study	Score	Study	Score
1	18	11.8	18	9.3	18	12.1	18	7.3
2	17	20	17	19.3	17	21.7	17	16.1
3	22	25	10	25	22	26.4	22	22
4	10	26.5	22	25	10	27.8	10	23.7
5	3	28.3	3	28.3	3	29.7	3	25.5
6	31	31.1	11	29.2	11	31.1	28	29.4
7	26b	32	26b	30.8	26b	32.5	11	29.8
8	11	32.1	28	34	28	35.3	30	30.4
9	28	33.9	12	35.3	31	35.3	26b	31.3
10	30	35.7	30	35.4	12	35.9	12	33.3
11	12	36.2	23	36.4	30	37.1	23	34
12	23	36.5	31	36.4	27	38.7	26a	34.5
13	26a	37.9	27	36.5	13	39.3	31	34.6
14	27	37.9	13	36.7	23	40.3	27	34.7
15	13	39.3	26a	36.8	26a	40.3	29	35.7
16	7	40.1	15	38.5	5	40.8	15	36.7
17	15	40.7	19	39.6	15	43.1	7	37.3
18	5	40.9	29	41.3	7	43.3	4	37.7
19	4	41.7	14	42	14	43.8	19	38
20	19	41.7	5	42.1	19	43.9	13	38.3
21	9	41.9	20	43.3	29	43.9	20	39
22	20	42	7	43.4	9	44.1	9	39.2
23	29	42.0	4	43.5	4	44.2	5	39.3
24	14	44.1	9	44.4	20	44.6	6	42.6
25	8	45.6	6	46.8	32	46.3	8	44.6
26	6	47.9	21	48.1	8	47.3	14	44.9
27	32	49.2	32	48.1	6	49.4	21	49
28	21	51.6	8	49.2	21	50	32	49.0
29	34	51.6	24	52.7	34	55.4	24	50.9
30	16	53.8	2	54.9	2	56.3	16	51.5
31	24	53.8	16	56.7	24	56.5	1	52.6
32	2	55.9	1	56.9	16	56.6	2	54
33	1	58.8	34	58.8	1	61.6	34	55.1
34	25	64.3	25	60	25	65.8	25	56.9
35	33	80.0	33	79.6	33	82.0	33	78.3

Table A2.4.3 Ranking of studies by mean score

Study	Mean Score	Rank
18	10.1	1
17	19.3	2
22	24.6	3
10	25.8	4
3	28.0	5
11	30.6	6
26b	31.7	7
28	33.2	8
31	34.4	9
30	34.7	10
12	35.2	11
23	36.8	12
27	37.0	13
26a	37.4	14
13	38.4	15
15	39.8	16
29	40.7	17
5	40.8	18
19	40.8	19
7	41.0	20
4	41.8	21
20	42.2	22
9	42.4	23
14	43.7	24
8	46.7	25
6	46.7	26
32	48.2	27
21	49.7	28
24	53.5	29
16	54.7	30
34	55.2	31
2	55.3	32
1	57.5	33
25	61.8	34
33	80.0	35

Table A2.4.4 Study ranking using the mean score: diagnostic accuracy studies

Study	Mean Score	Rank
18	10.1	1
17	19.3	2
22	24.6	3
10	25.8	4
3	28.0	5
11	30.6	6
23	36.8	12
27	37.0	13
13	38.4	15
14	43.7	24
2	55.3	32
All studies	**31.8**	

Table A2.4.5 Study ranking using the mean score: diagnostic performance studies

Study	Mean Score	Rank
28	33.2	8
30	34.7	10
12	35.2	11
26a	37.4	14
15	39.8	16
29	40.7	17
19	40.8	19
7	41.0	20
9	42.4	23
8	46.7	25
32	48.2	27
21	49.7	28
34	55.2	31
25	61.8	34
33	80.0	35
All studies	**45.8**	

Table A2.4.6 Study ranking using the mean score: diagnostic outcome studies

Study	Mean Score	Rank
31	34.4	9
26b	31.7	7
5	40.8	18
4	41.8	21
20	42.2	22
6	46.7	26
24	53.5	29
16	54.7	30
1	57.5	33
All studies	**44.8**	

Annex 2.5 Exploring sources of heterogeneity

Table A2.5.1 Raw data for the studies included in the regression

Covariate	Ref. No.											
	28	31	30	12	5	4	9	8	32	21	16	34
B1	A	A	A	A	A	A	A	C	A	A	A	A
B2	A	A	A	A	A	A	A	A	A	A	A	A
B3	A	C	A	A	C	A	A	A	C	C	A	C
B4	A	A	A	A	A	A	A	A	A	C	A	A
B5	A	C	A	A	A	A	A	A	C	A	A	C
C1	A	A	A	A	A	A	A	A	A	A	A	C
C2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
D1	A	A	A	A	C	C	A	A	A	C	C	C
D2	A	N/A	N/A	A	C	C	A	A	N/A	C	C	C
D3	C	A	A	A	A	A	C	C	C	A	B	C
E1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
E2	C	A	A	A	A	A	A	A	A	C	A	A
E3	A	C	A	A	A	A	B	B	A	A	C	C
E4	C	N/A	N/A	C	B	C	C	C	N/A	C	C	N/A
E5	C	N/A	N/A	C	C	C	C	C	C	C	A	N/A
E6	C	C	C	N/A	C	C	C	C	C	C	C	C
E7	N/A	C	C	N/A	C	C	C	C	C	C	C	C
F1	A	A	A	C	A	A	A	A	A	A	A	A
F2	C	A	A	A	A	C	A	A	C	A	B	A
F3	A	N/A	N/A	B	N/A	B	A	B	C	A	B	N/A
F4	N/A	A	N/A	N/A	N/A	N/A	N/A	A	C	N/A	A	N/A
F5	A	A	N/A	A	A	A	A	A	A	A	A	A
F6	C	A	C	C	A	C	C	B	A	C	B	C
F7	C	C	C	C	C	C	C	C	C	C	C	A
F8	C	A	C	A	C	C	B	C	A	A	C	C
F9	N/A	A	N/A	N/A	N/A	A	N/A	N/A	N/A	N/A	A	N/A
F10	N/A	A	N/A	N/A	A	A	N/A	N/A	N/A	N/A	A	N/A
F11	N/A	C	N/A	N/A	C	B	N/A	N/A	N/A	N/A	C	N/A
G1	A	C	C	C	C	C	C	C	C	C	C	C
H1	A	A	C	A	A	A	A	A	A	A	A	A
H2	A	A	A	A	A	A	C	A	A	A	C	A
H3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H4	C	A3	A3	A3	A2	A3	C	C	C	A2	A3	A3
H5	A3	A6	A4	C	A6	A4	C	C	C	C	C	A5
H6	N/A	N/A	C	N/A	N/A	N/A	N/A	C	N/A	N/A	C	N/A
I1	A	N/A	A	A	A	A	A	A	A	C	C	C
I2	A	A	A	A	A	A	A	A	A	A	A	A
I3	A	N/A	N/A	N/A	A	A	N/A	N/A	N/A	N/A	A	A
Training?	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Skeleton?	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
A&E?	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No
Study type	DP	DO	DP	DP	DO	DO	DP	DP	DP	DP	DO	DP
Grey Literature?	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Quality	B	B	B	B	B	B	B	B	B	C	C	C

Table A2.5.2 Variables included in the regression and their values

Covariate	0	1
BIASES		
Film selection		
Is centripetal bias present? (B1)	n	y
Is population bias present? (B3)	n	y
Is eligibility criteria bias present? (B4)	n	y
Is film selection bias present (B5)	n	y
Observer selection		
Is observer cohort bias present? (C1)	n	y
Application of the standard		
Is verification bias present? (D1)	n	y
Is work-up bias present? (D2)	n	y
Is incorporation bias present? (D3)	n	y
Measurement of results		
Is indeterminate bias present? (E2)	n	y
Are there any patients lost to follow-up? (E3)	n	y
Any attempt to assess inter-observer variability? (E5)	n	y
Independence of interpretations		
Is observer review bias present? (F1)	n	y
Is reference standard review bias present? (F2)	n	y
Is observer bias present? (F3)	n	y
Is observer comparator bias present? (F4)	n	y
Is arbiter review bias present? (F6)	n	y
Is arbiter bias present? (F7)	n	y
Is film access bias present? (F8)	n	y
METHODOLOGICAL FACTORS		
Was sample size calculated? (G1)	n	y
Was a report adequately defined? (H1)	n	y
Assessed in the correct diagnostic sequence? (H2)	n	y
Was there a valid reference standard? (H4)	n	y
Was there a valid arbiter (H5)	n	y
Appropriate subgroup analyses? (I1)	n	y
OTHER FACTORS		
Had any training?	n	y
Type of body areas included	Any body area	Skeletal only
Type of patients included	Any patient type	A&E only
Type of study design used	Performance	Outcome
Was the study grey literature?	n	y
Quality of the study?	C	B

Table A2.5.3 Data for studies in regression

Covariate	**Ref. No.**											
	28	**31**	**30**	**12**	**5**	**4**	**9**	**8**	**32**	**21**	**16**	**34**
B1	0	0	0	0	0	0	0	1	0	0	0	0
B3	0	1	0	0	1	0	0	0	1	1	0	1
B4	0	0	0	0	0	0	0	0	0	1	0	0
B5	0	1	0	0	0	0	0	0	1	0	0	1
C1	0	0	0	0	0	0	0	0	0	0	0	1
D1	0	0	0	0	1	1	0	0	0	1	1	1
D2	0	N/A	N/A	0	1	1	0	0	N/A	1	1	1
D3	1	0	0	0	0	0	1	1	1	0	1	1
E2	1	0	0	0	0	0	0	0	0	1	0	0
E3	0	1	0	0	0	0	1	1	0	0	1	1
E5	1	N/A	N/A	1	1	1	1	1	1	1	0	N/A
F1	0	0	0	1	0	0	0	0	0	0	0	0
F2	1	0	0	0	0	1	0	0	1	0	1	0
F3	0	N/A	N/A	1	N/A	1	0	1	1	0	1	N/A
F4	N/A	0	N/A	N/A	N/A	N/A	N/A	0	1	N/A	0	N/A
F6	1	0	1	1	0	1	1	1	0	1	1	1
F7	1	1	1	1	1	1	1	1	1	1	1	0
F8	1	0	1	0	1	1	1	1	0	0	1	1
G1	1	0	0	0	0	0	0	0	0	0	0	0
H1	1	1	0	1	1	1	1	1	1	1	1	1
H2	1	1	1	1	1	1	0	1	1	1	0	1
H4	0	1	1	1	1	1	0	0	0	1	1	1
H5	1	1	1	0	1	1	0	0	0	0	0	1
I1	1	N/A	1	1	1	1	1	1	1	0	0	0
Training?	1	0	1	1	1	1	1	1	0	1	1	1
Skeleton?	1	1	1	1	0	0	1	1	1	1	1	1
A&E?	1	1	1	1	1	1	0	0	1	1	1	0
Study type	0	1	0	0	1	1	0	0	0	0	1	0
Grey Literature?	1	1	0	1	0	1	1	1	1	1	0	1
Quality?	1	1	1	1	1	1	1	1	1	0	0	0

Table A2.5.4 Results of separately regressing D on each explanatory variable

Variable	α		β_0		β_1			
	Estimate	*P-Value*	*Estimate*	*P-Value*	*Estimate*	*SE*	*95% CI*	*P-value*
B1	5.258	0.000	-0.687	0.326	0.125	0.884	-1.9, 2.1	0.891
B3	5.501	0.000	-0.532	0.443	-0.281	0.867	-2.2, 1.7	0.753
B4	5.343	0.000	-0.632	0.360	0.037	1.285	-2.9, 2.9	0.978
B5	5.424	0.000	-0.584	0.352	-0.396	1.116	-2.9, 2.1	0.731
C1	5.405	0.000	-0.607	0.316	-0.626	1.174	-3.3, 2.0	0.607
D1	5.860	0.000	-0.427	0.430	-0.904	0.528	-2.1, 0.3	0.121
D2	5.761	0.001	-0.484	0.508	-0.862	0.647	-2.4, 0.7	0.231
D3	5.353	0.000	-0.638	0.298	-0.003	0.563	-1.3, 1.2	0.955
E2	4.691	0.000	-0.991	0.053	1.672	0.611	0.3, 3.0	0.023
E3	5.858	0.000	-0.564	0.293	-0.808	0.479	-1.9, 0.3	0.126
E5	4.404	0.004	-0.607	0.351	1.183	0.743	-0.6, 3.0	0.162
F1	5.323	0.000	-0.668	0.301	-0.090	0.648	-1.6, 1.4	0.982
F2	5.279	0.000	-0.631	0.299	0.254	0.625	-1.2, 1.7	0.694
F3	5.598	0.003	-0.542	0.483	-0.751	0.710	-2.6, 1.1	0.339
F4	6.079	0.072	0.745	0.499	-0.062	1.837	-23.4, 23.3	0.979
F6	5.508	0.005	-0.659	0.294	-0.207	1.485	-3.6, 3.2	0.892
F7	4.779	0.004	-0.607	0.316	0.626	1.174	-2.0, 3.3	0.607
F8	5.308	0.000	-0.640	0.296	0.036	0.574	-1.3, 1.3	0.951
G1	5.230	0.000	-0.556	0.101	2.636	0.544	1.4, 3.9	0.001
H1	7.021	0.023	-0.677	0.261	-1.752	2.569	-7.6, 4.1	0.512
H2	4.312	0.001	-0.981	0.101	0.930	0.539	-0.3, 2.1	0.118
H4	5.765	0.000	-0.645	0.244	-0.739	0.498	-1.9, 0.4	0.172
H5	5.103	0.000	-0.652	0.206	1.194	0.588	-0.1, 2.5	0.073
I1	4.987	0.000	-0.278	0.639	1.067	0.580	-0.3, 2.4	0.103
Training?	6.151	0.035	-0.724	0.274	-0.930	2.691	-7.0, 5.1	0.738
Skeleton?	5.528	0.013	-0.627	0.314	-0.188	1.562	-3.7, 3.3	0.907
A&E?	5.216	0.000	-0.632	0.299	0.200	0.559	-1.1, 1.5	0.729
Study type	5.492	0.000	-0.659	0.233	-0.931	0.614	-2.3, 0.5	0.164
Grey Literature?	4.661	0.000	-0.610	0.274	0.875	0.625	-0.5, 2.3	0.195
Pre-2000?	6.296	0.000	-0.79	0.136	-1.332	0.649	-2.8, 0.1	0.070
Quality?	4.952	0.000	-0.316	0.553	1.049	0.542	-0.2, 2.3	0.085

Annex 2.6 Heterogeneity in presence of biases and methodological standards

Table A2.6.1 Presence of bias in the different study designs

Criterion	Diagnostic Accuracy				Diagnostic Performance				Diagnostic Outcome			
	A	B	C	NA	A	B	C	NA	A	B	C	NA
A1	11	0	0	0	**	**	**	**	**	**	**	**
A2	8	2	1	0	**	**	**	**	**	**	**	**
B1	**	**	**	**	14	0	1	0	9	0	0	0
B2	**	**	**	**	15	0	0	0	9	0	0	0
B3	**	**	**	**	11	0	4	0	5	0	4	0
B4	**	**	**	**	12	1	2	0	8	0	1	0
B5	**	**	**	**	11	1	3	0	4	0	5	0
C1	10	0	1	0	13	1	1	0	9	0	0	0
C2	3	0	0	8	**	**	**	**	**	**	**	**
D1	11	0	0	0	12	1	2	0	6	0	3	0
D2	10	0	0	1	9	0	2	4	5	0	3	1
D3	11	0	0	0	9	1	5	0	8	1	0	0
E1	**	**	**	**	**	**	**	**	**	**	**	**
E2	10	1	0	0	9	1	2	3	7	0	2	0
E3	**	**	**	**	10	4	1	0	5	1	2	1
E4	0	0	9	2	0	0	10	5	0	1	7	1
E5	1	0	10	0	0	0	10	5	1	0	7	1
E6	0	0	11	0	0	0	14	1	0	0	9	0
E7	1	0	9	1	0	0	13	2	0	0	9	0
F1	11	0	0	0	14	0	1	0	9	0	0	0
F2	11	0	0	0	6	0	9	0	3	1	5	0
F3	9	1	1	0	6	3	1	5	1	3	0	6
F4	3	0	0	8	1	1	1	12	2	0	0	7
F5	10	1	0	0	13	1	0	1	9	0	0	0
F6	9	1	1	0	1	4	10	0	3	2	4	0
F7	0	3	8	0	1	1	13	0	0	2	7	0
F8	2	1	8	0	3	3	9	0	1	2	6	0
F9	6	0	0	5	**	**	**	**	4	1	1	3
F10	6	0	0	5	**	**	**	**	9	0	0	0
F11	1	0	5	5	**	**	**	**	0	2	7	0

A = Done, B = Not Clear, C = Not Done, NA = Not applicable

Table A2.6.2 Methodological standards met for the different study designs

Criterion	Diagnostic Accuracy			Diagnostic Performance			Diagnostic Outcome		
	DONE	NOT DONE	NA	DONE	NOT DONE	NA	DONE	NOT DONE	NA
G1	0	11	0	2	13	0	0	9	0
H1	11	0	0	12	3	0	8	1	0
H2	**	**	**	14	1	0	8	1	0
H3	**	**	**	**	**	**	1	1	8
H4	11	0	0	8	7	0	8	1	0
H5	9	2	0	4	11	0	7	2	0
H6	2	2	7	0	5	10	2	0	7
I1	4	4	3	7	7	1	3	5	1
I2	4	5	2	11	4	0	6	3	0
I3	2	2	7	4	1	10	6	2	1

Table A2.6.3 Presence of bias in all studies

Criterion	All studies				Total	%
	A	B	C	NA		
A1	11	0	0	0	0/11	0.0
A2	8	2	1	0	3/11	27.3
B1	23	0	1	0	1/24	4.2
B2	24	0	0	0	0/24	0.0
B3	16	0	8	0	8/24	33.3
B4	20	1	3	0	4/24	16.7
B5	15	1	8	0	9/24	37.5
C1	22	1	1	0	2/24	8.3
C2	3	0	0	8	0/3	0.0
D1	29	1	5	0	6/35	17.1
D2	24	0	5	6	5/29	17.2
D3	28	2	5	0	7/35	20.0
E1	**	**	**	**	**	**
E2	26	2	4	3	6/32	18.8
E3	15	5	3	1	8/23	34.8
E4	0	1	26	8	27/27	100.0
E5	2	0	27	6	27/29	93.1
E6	0	0	34	1	34/34	100.0
E7	1	0	31	3	31/32	96.9
F1	34	0	1	0	1/35	2.9
F2	11	10	14	0	24/35	68.6
F3	16	7	2	11	9/25	36.0
F4	6	1	1	27	2/8	25.0
F5	32	2	0	1	2/34	5.9
F6	13	8	15	0	23/36	63.9
F7	1	6	28	0	34/35	97.1
F8	6	6	23	0	29/35	82.9
F9	10	1	1	8	2/12	16.7
F10	15	0	0	5	0/15	0.0
F11	1	12	2	5	14/15	93.3

Table A2.6.4 Methodological standards met for all studies

Criterion	All Studies			Total	%
	DONE	NOT DONE	NA		
G1	2	33	0	2/35	5.7
H1	31	4	0	31/35	88.6
H2	22	2	0	22/24	91.7
H3	1	1	8	1/2	50.0
H4	27	8	0	27/35	77.1
H5	20	15	0	20/35	57.1
H6	4	7	24	4/11	36.4
I1	14	16	5	14/30	46.7
I2	21	12	2	21/33	63.6
I3	12	5	18	12/17	70.6

Annex 4 Proformas

Proforma A4.1 Form used to record reference standard report

Age		D.O.E		Examination		XR No.	
Sex		D.O.R				Ref No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	
Clinical Details				**Report**			

Proforma A4.2 Form used by arbiter to judge concordance between pairs of reports

A&E/GP					
Age		Examination required		Hospital No.	
				Reference No.	
Sex					
Walking		Ambulance S1/2		Trolley	
Chair		Stretcher		Escort Yes/No	
Date of Examination					

Clinical Details

Report	**Report**
Criteria: Please tick in the appropriate boxes	
Definitely normal [ ] Probably normal [ ] Possibly abnormal [ ] Probably abnormal [ ] Definitely abnormal [ ]	Definitely normal [ ] Probably normal [ ] Possibly abnormal [ ] Probably abnormal [ ] Definitely abnormal [ ]
Normal [ ] Abnormal [ ] Equivocal [ ]	Normal [ ] Abnormal [ ] Equivocal [ ]
Agree: [ ] **Disagree:** [ ]	

Proforma A4.3 Form used to record A&E patient management details

<table>
<tr><td>Age</td><td></td><td>D.O.E</td><td></td><td colspan="2" rowspan="2">Examination required</td><td>Hospital No.</td><td></td></tr>
<tr><td>Sex</td><td></td><td>D.O.R</td><td></td><td>Reference No.</td><td></td></tr>
<tr><td>A&E/GP</td><td></td><td colspan="2" rowspan="3"></td><td>Walking</td><td></td><td>Ambulance S1/2</td><td></td></tr>
<tr><td>Trust</td><td></td><td>Chair</td><td></td><td>Stretcher</td><td></td></tr>
<tr><td colspan="2"></td><td>Trolley</td><td></td><td>Escort Yes/No</td><td></td></tr>
<tr><td colspan="8">Clinical details (block capitals)</td></tr>
<tr><td colspan="8">Initial Management (block capitals)</td></tr>
<tr><td colspan="8">Patient recalled and observed management
(during a one month period from the time of the initial examination)</td></tr>
</table>

Proforma A4.4 Form used to record affect of reference standard report on A&E patient management and outcome

Age		D.O.E		Examination required		Hospital No.	
Sex		D.O.R				Reference No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	

Clinical details (block capitals)

Initial Management (block capitals)

Report

Patient recalled and expected management

Proforma A4.5 Form used to judge whether differences in A&E patient management may affect patient outcome

Age		D.O.E		Examination required		Hospital No.	
Sex		D.O.R				Reference No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	

Clinical details (block capitals)

Initial Management (block capitals)

Report	**Report**

Patient recalled and management	**Patient recalled and management**

1. No difference in patient management (treatment or advice):

2. A clinically unimportant difference in patient management:

3. A clinically important difference in patient management:

Proforma A4.6 Form used to record affect of report on GP patient management

Age		D.O.E		Examination required		Hospital No.	
Sex		D.O.R				Reference No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	
Clinical details							
Report							
Expected management							

Proforma A4.7 Form used to judge whether differences in GP patient management may affect patient outcome

Age		D.O.E		Examination required		Hospital No.	
Sex		D.O.R				Reference No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	
Clinical details							
Report				**Report**			
Expected management				**Expected management**			
1. No difference in the expected patient management (treatment or advice):							
2. A clinically unimportant difference to the expected patient management:							
3. A clinically important difference to the expected patient management:							

Annex 5 Proformas and Additional marking criteria

Proformas

Proforma A5.1 Form used to record reports

Age		D.O.E		Examination		XR No.	
Sex		D.O.R				Ref No.	
A&E/GP				Walking		Ambulance S1/2	
Trust				Chair		Stretcher	
				Trolley		Escort Yes/No	
Clinical Details				**Report**			

Proforma A5.2 Form used by arbiter to judge concordance between pairs of reports

PROFORMA					
A&E/GP					
Age		Examination required		Hospital No.	
Sex				Reference No.	
Walking		Ambulance S1/2		Trolley	
Chair		Stretcher		Escort Yes/No	
Date of Examination					

Clinical Details

Report	**Report**

Criteria: Please tick in the appropriate boxes

Definitely normal [ ]
Probably normal [ ]
Possibly abnormal [ ]
Probably abnormal [ ]
Definitely abnormal [ ]

Normal [ ]
Abnormal [ ]
Equivocal [ ]

Definitely normal [ ]
Probably normal [ ]
Possibly abnormal [ ]
Probably abnormal [ ]
Definitely abnormal [ ]

Normal [ ]
Abnormal [ ]
Equivocal [ ]

Agree: [ ]
Disagree: [ ]

Proforma A5.3 Pre-report questionnaire for recording diagnosis and management

PROFORMA						
Age		D.O.E.		Examination required		Hospital No.
Sex		D.O.R.				Reference No.
A&E/GP				Walking		Ambulance S1/2
Trust				Chair		Stretcher
				Trolley		Escort Yes/No

Clinical details

PRE-REPORT QUESTIONNAIRE

These questions are to be answered before the report is available:

1. What do you believe to be the patient's diagnosis?

2. On a scale of 0 per cent (uncertain) to 100 per cent (certain), how certain are you of the diagnosis?

3. What is the proposed treatment plan for the patient?

4. On a scale of 0 per cent (uncertain) to 100 per cent (certain), how certain are you that the treatment plan is appropriate?

Proforma A5.4 Post-report questionnaire for recording diagnosis and management

PROFORMA						
Age		D.O.E		Examination required		Hospital No.
Sex		D.O.R				Reference No.
A&E/GP				Walking		Ambulance S1/2
Trust				Chair		Stretcher
				Trolley		Escort Yes/No

Clinical details

Report

POST-REPORT QUESTIONNAIRE

These questions are to be answered when the report is available:

1. What do you believe to be the patient's diagnosis?

2. On a scale of 0 per cent (uncertain) to 100 per cent (certain), how certain are you of the diagnosis?

3. What is the proposed treatment plan for the patient?

4. On a scale of 0 per cent (uncertain) to 100 per cent (certain), how certain are you that the treatment plan is appropriate?

5. Please list below other diagnostic tests (e.g. bone scan, CT, Dexta scan) that you are considering to use?

Proforma A5.5 Form used to judge whether differences in management may affect patient outcome

PROFORMA						
Age		D.O.E		Examination required		Hospital No.
Sex		D.O.R				Reference No.
A&E/GP				Walking		Ambulance S1/2
Trust				Chair		Stretcher
				Trolley		Escort Yes/No

Clinical details

Report	**Report**
Expected management	**Expected management**

1. No difference in the expected patient management (treatment or advice):
2. A clinically unimportant difference to the expected patient management:
3. A clinically important difference to the expected patient management:

Additional marking criteria

After the original marking criteria had been applied it was discovered that sometimes when both professional groups reported the same radiograph a CSRs report was classified as a true negative (TN) but a radiologist report as a true positive (TP). If one report was classified as correctly normal (TN) the other report could not be classified as correctly abnormal (TP). The following guidelines were given to the external arbiter to resolve this problem.

TN vs TP

If both reports were made on the same radiograph one can't be classified as correctly normal (TN) and the other report as correctly abnormal (TP). For the problem of TN vs TP do not change your original decision that both reports agree with the standard but please record whether they are TN or TP.

TN vs FN or TP vs FP

If both reports are made on the same radiograph one can't be classified as correctly normal (TN) and the other as incorrectly normal (FN). Similarly, if both reports are made on the same radiograph one can't be classified as correctly abnormal (TP) and the other incorrectly abnormal (FP).

a) TN vs FN - change the TN to be a TP; or change the FN to a FP. Choose the option that most closely reflects what both reports are in comparison to the standard report.

b) TP vs FP - change the TP to a TN, or change the FP to a FN.

FN vs FP

If both reports are made on the same radiograph one can't be classified as incorrectly normal (FN) and the other as incorrectly abnormal (FP). Similar to TN vs TP re-classify the reports so they are FN or FP.

Annex 6 Cost Questionnaire and Scenario Analysis

Cost Questionnaire

DIRECT COSTS

A SETTING-UP COSTS

1. Fixed

CAPITAL COSTS

<u>Equipment</u>

1. What was the cost of the following equipment at the time of installation (i.e. £x/199x), which was provided for the CSR to use when reporting? Please record the number of items that were purchased and the cost per item:

- Personal computer
- Printer
- viewing boxes
- chair
- worktops

2. What was the cost of installing the following equipment (i.e. £x/199x)? Please record the number of items that were purchased and the cost of installation per item:

- Personal computer
- Printer
- viewing boxes
- chair
- worktops

3. What was the cost of adding a partition to the room that was converted into the CSRs reporting room and the photocopier/fax/storage room (£x/199x)?

TRAINING COSTS [1]

4. What was the cost (£x/199x) of the MSc course at Leeds?

5. What were the CSR traveling expenses when attending the MSc course?

6. How many hours were the CSRs not available for work because of the course from September 1992 to 31/1/93 and then for each year afterwards i.e. 1/2/93 - 31/1/94 etc?

STAFFING COSTS

1. What was the cost of employing the basic radiographer permanently from January 1993 onwards?

2. Has the introduction of CSR reporting resulted in any extra medico-legal costs?

B COSTS OF RUNNING THE SERVICE

Fixed costs

Staffing costs

1. The CSRs were upgraded from Senior II to Senior I from January 1993 onwards. What was their weekly salary in September 1992 and January 1993? When and what were their other salary increases until the end of January 1997?

2. What was the weekly average salary of a consultant radiologist in September 1992 and January 1993? When and what were their other salary increases until the end of January 1997?

3. What is the weekly average salary of the secretaries for each year (i.e. 1/2/93 - 31/1/94 etc.) during the single time period (1/2/93 - 31/1/98)? State when and what changes were made to their salary.

---

[1] Training began in Sept 92 and ended in Dec 1994. The radiographers started to report on normal films in Feb 95 and also abnormal films as well from Feb 96.

Variable costs (time costs)

1. What proportion of radiologists and CSRs time is spent reporting (difference in times for different areas of the body)?

2. The CSRs have to type their own reports into the computer system, whereas the radiologists have the help of the secretaries. How much time do the secretaries spend on typing reports of A&E examinations for the radiologists?

3. No specific time was allocated to the assessment of CSR reporting. However the CSRs are expected to undertake a review of their work at least every 12 months.

a. When have the CSRs done this? How much time was involved (audit)?

b. The CSRs keep a record of the radiographs they report on for quality assurance. How many hours do they spend doing this?

4. State the costs associated with the learning effects of a new technology *:

a. As the radiographers have become more experienced, they may not seek the advice of the radiologists as much as they did originally.

(i) How often did the CSRs asked radiologists for their advice and how this changed over time?

(ii) Identify the probable amount of time it takes to seek the advice of a radiologist.

b. Are there any other changes that may have occurred as a consequence of the CSRs gaining more experience and confidence?

5. State whether any adjustments need to be made for protocol driven costs.

a. This study may possibly underestimate the cost of CSR reporting because it was not possible to time how long it took radiographers to report originally.

b. Can you think of any other ways in which this study may over/under estimate the cost of CSR or consultant radiologist reporting?

c. Overhead costs

Consumables

1. This includes bulbs to the viewing boxes, heat, the electrical power used, printed labels, reports and print ribbons. These consumables are part of the reporting process whether it is radiographers or not. Only if there is a large (and unlikely) difference in the time it takes radiographers to report A&E examinations compared with the radiologists will there be any difference in cost. It doesn't matter whether they are used by a radiographer or a radiologist, with the exception of the bulbs to the viewing boxes and electrical power depending on whether there is a. For the moment please simply provide me with the current relevant cost/quantity for the following items e.g. cost/100 print ribbons:

- Bulbs to the viewing boxes (cost/quantity)
- Heat (cost/hr)
- Electrical power (cost/hr)
- Printed labels (cost/quantity)
- Paper used for the reports (cost/quantity)
- Print ribbons (cost/quantity)

2. The radiologists make use of the secretaries and the CSRs do not. What are the cost of tapes and tape recorders and other relevant equipment? What are the associated costs for each year of the time period 1/2/93 - 31/1/98?

3. Rent/floor space

i) a) What is the floor space (m^2) of the <u>radiographers reporting room</u>?

b) Is this room used for anything else other than for its original purpose (e.g. photocopying) and for conference and clinical meetings?

ii) a) What is the floor space (m^2) of the individual <u>radiologists offices</u>?

iii) a) What is the floor space (m^2) of the <u>hot reporting room</u> used by the radiologists?

b) Is this room used for anything else?

Scenario Analysis

CSR or radiologist reporting A&E radiographs of the appendicular skeleton

WORSE CASE SCENARIO

This scenario presents assumptions against CSR reporting:

- All capital costs are additional to the cost of CSR reporting
- The time period over which training is annuitized is 10 years
- Capital and training costs use 7 per cent interest rate for annual equivalent cost (AEC)
- All overhead costs of CSR reporting room are additional costs to CSR reporting
- No overhead costs are additional to radiologist reporting
- Radiographer minute is 1.4

CSR reporting costs:

Capital costs at 7 per cent AEC	= £892.41
Training costs annuitized over ten years at 7 per cent AEC	= £1056.28
Overhead costs for CSR reporting room	= £3240.85
Sub Total	= £5189.54

- CSR hourly rate = £18.90/hr
- Time to report and type report = 65.88 * 1.4 = 92.23 seconds
- Cost per radiograph reported = 92.23 / 3600 = 0.026 * £18.90 = £0.49
- The annual cost is £0.49 * 9713 examinations = £4759.37
- Total annual cost of CSR reporting = £5189.54 + £4759.37 = £9948.91
- Average cost = £9948.91 / 9713 = £1.02/radiograph reported

Radiologist reporting costs:

- Radiologist hourly rate = £33.66/hr; Time to report = 42.50 seconds
- Cost per radiograph reported = 42.50 / 3600 = 0.012 * £33.66 = £0.40
- Annual cost is £0.40 * 9713 examinations = £3885.20
- Secretary hourly rate = £11.50/hr
- Time to type report is same as radiologist time to report i.e. 42.50 seconds.
- Cost per radiograph typed by a secretary is = 42.50 / 3600 = 0.012 * £11.50 = £0.14.
- Annual cost of secretary typing is £0.14 * 9713 examinations = £1359.82
- Total annual cost of radiologist reporting = £3885.20 + £1359.82 = £5425.02
- Average cost = £5425.02 / 9713 = £0.54/radiograph reported

BEST CASE SCENARIO

This scenario presents assumptions in favour of CSR reporting:

- Capital costs – 50 per cent attributed to CSR reporting; 25 per cent attributed to radiologist reporting; (25 per cent attributed to general use)
- The time period over which training is annuitized is 20 years
- For capital and training costs use 5 per cent interest rate for AEC
- Overhead costs of CSR reporting room – 50 per cent are attributed to CSR reporting; 25 per cent attributed to radiologist reporting; (25 per cent attributed to general use)
- Overhead costs of radiologist and secretary office – 10 per cent attributed to radiologist reporting
- No secretarial costs if they type both CSR and radiologist dictated reports
- Radiographer minute is equivalent to the radiologist
- Radiographer reporting time is equivalent to radiologist

CSR reporting costs:

50 per cent of the capital costs at 5 per cent AEC	= £426.39
Training costs annuitized over 20 years at 5 per cent AEC	= £605.97
50 per cent of overhead costs for CSR reporting room	= £1620.43
Sub Total	= £2652.79

- CSR hourly rate = £18.90/hr
- Time to report and type report = 42.50 seconds
- Cost per radiograph reported by CSR = 42.50 / 3600 = 0.012 * £18.90 = £0.23
- The annual cost is £0.23 * 9713 examinations = £2233.99
- Total annual cost of CSR reporting is £2233.99 + £2652.79 = £4886.78
- Average cost = £4886.78 / 9713 = £0.50/radiograph reported

Radiologist reporting costs:

25 per cent of the capital costs at 5 per cent AEC	= £213.20
25 per cent of overhead costs for CSR reporting room	= £810.21
10 per cent of overhead costs of radiologist / secretary office	= £4506.30
Sub Total	= £5529.71

- The total annual cost of radiologist reporting is again £3885.20
- Total cost of radiologist reporting is £3885.20 + £5529.71 = £9414.91
- Average cost = £9414.91 / 9713 = £0.97/radiograph reported

BASE CASE SCENARIO

This is the most realistic scenario and assumes the following:

- 50 per cent of the capital costs are additional to the cost of CSR reporting
- The time period over which training is annuitized is 15 years
- For capital and training costs use 6 per cent interest rate for AEC
- 50 per cent of overhead costs of CSR reporting room are attributed to CSR reporting
- 5 per cent of overhead costs of radiologist and secretary floor space are attributed to radiologist reporting
- Typing reports by secretaries is an additional cost to radiologist reporting
- Radiographer minute is 1.3

CSR reporting costs:

50 per cent of the capital costs at 6 per cent AEC	= £374.68
Training costs annuitized over 15 years at 6 per cent AEC	= £771.07
50 per cent of overhead costs for CSR reporting room	= £1620.43
Sub Total	= £2766.18

- CSR hourly rate = £18.90/hr
- Time to report and type report = 65.88 * 1.3 = 85.64 seconds
- Cost per radiograph reported by CSR = 85.64 / 3600 = 0.024 * £18.90 = £0.45
- The annual cost is £0.45 * 9713 examinations = £4370.85
- In addition there are the other annual costs of £2766.18
- Total annual cost of CSR reporting = £7137.03
- Average cost = £7137.03 / 9713 = £0.73/radiograph reported

The following describes the base case cost of CSR reporting if the secretaries typed their reports:

- CSR time to report assuming the secretary types = 47.13 * 1.3 = 61.27 seconds
- Cost per radiograph reported by CSR = 61.27 / 3600 = 0.017 * £18.90 = £0.32
- Total annual cost is £0.32 * 9713 examinations = £3108.16
- Assuming a secretary would take the same time to type as the CSR = 18.75 / 3600 = 0.005 * £11.50 = £0.06
- The annual cost is £0.06 * 9713 examinations = £582.78
- Total cost of CSR reporting is £2766.18 + £3108.16 + £582.78 = £6457.12
- Average cost = £6427.12 / 9713 = £0.66/radiograph reported

Radiologist reporting costs:

- The total annual cost is again £3885.20
- 5 per cent of overhead costs of radiologist and secretary office space = £2253.15
- Secretary hourly rate = £11.50/hr
- Time to type report is same as for when radiologist reports, which is 42.50 seconds. Cost per radiograph typed by a secretary is = 42.50 / 3600 = 0.012 * £11.50 = £0.14.
- Annual cost of secretary typing is £0.14 * 9713 examinations = £1359.82
- Total cost of reporting is £3885.20 + £2253.15 + £1359.82 = £7498.17
- Average cost = £7498.17 / 9713 = £0.77/radiograph reported

CSR and radiologist reporting of all A&E radiographs

This is the base case or most realistic scenario and assumes the following:

- 75 per cent of the capital costs are additional to the cost of CSR reporting
- The time period over which training is annuitized is 15 years
- For capital and training costs use 6 per cent interest rate for AEC
- 75 per cent of overhead costs of CSR reporting room are attributed to CSR reporting
- 5 per cent of overhead costs of radiologist and secretary floor space are attributed to radiologist reporting
- Typing reports by secretaries is an additional cost to radiologist reporting
- Radiographer minute is 1.3

CSR reporting costs:

75 per cent of the capital costs at 6 per cent AEC	= £562.02
Training costs annuitized over 15 years at 6 per cent AEC	= £771.07
75 per cent of overhead costs for CSR reporting room	= £2430.64
Sub Total	= £3763.73

- CSR hourly rate = £18.90/hr
- CSR time to interpret, record and type = 102.90 * 1.3 = 133.77 seconds
- Cost per radiograph reported by CSR = 133.77 / 3600 = 0.037 * £18.90 = £0.70
- The annual cost is £0.70 * 20000 examinations = £14000
- In addition there are the other annual costs of £3763.73
- Total annual cost of CSR reporting = £17763.73
- Average cost = £17763.73 / 20000 = £0.89/radiograph reported

The following describes the base case cost of CSR reporting if the secretaries typed their reports:

- Assuming a secretary would take the same time to type as before = 18.75 / 3600 = 0.005 * £11.50 = £0.06
- The annual cost is £0.06 * 20000 examinations = £1200
- CSR time to interpret and record = 84.15 * 1.3 = 109.40 seconds
- Cost per radiograph reported by CSR = 109.40 / 3600 = 0.030 * £18.90 = £0.57
- The annual cost is £0.57 * 20000 examinations = £11400
- Total cost of CSR reporting is £3763.73 + £1200 + £11400 = £16363.73
- Average cost = £163673.73 / 20000 = £0.82/radiograph reported

Radiologist reporting costs:

- 5 per cent of overhead costs of radiologist and secretary office space = £2253.15
- Radiologist hourly rate = £33.66/hr
- Time to report = 51.96 seconds
- Cost per radiograph reported = 51.96 / 3600 = 0.014 * £33.66 = £0.47
- Annual cost is £0.47 * 20000 examinations = £9400
- Secretary hourly rate = £11.50/hr
- Time to type report is same as for when radiologist reports, which is 51.96 seconds. Cost per radiograph typed by a secretary is = 51.96 / 3600 = 0.014 * £11.50 = £0.16
- Annual cost of secretary typing is £0.16 * 20000 examinations = £3200
- Total cost of reporting is £2253.15 + £9400 + £3200 = £14853.15
- Average cost = £14853.15 / 20000 = £0.74/radiograph reported

List of abbreviations

A&E	Accident and Emergency
Acc	Accuracy
Append	Appendicular
BJR	British Journal of Radiology
CCTR	Cochrane Controlled Trials Register
CDSR	Cochrane Database of Systematic Reviews
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CO	Casualty Officer
CoR	College of Radiographers
CRD	Centre for Reviews and Dissemination
CSR	Clinical Specialist Radiographers
CT	Computed Tomography
DARE	Database of Abstracts of Reviews of Effectiveness
DH	Department of Health
DOR	Diagnostic Odds Ratio
EMBASE	*Excerpta Medica* online
ENP	Emergency Nurse Practitioner
FN	False Negative
FP	False Positive
GP	General Practitioner
HTA	Heath Technology Assessment
IP	In-patient
LR	Likelihood Ratios
MEDLINE	*Index Medicus* online
MRI	Magnetic Resonance Imaging
NHS EED	NHS Economic Evaluation database
NHS	National Health Service
NM	Nuclear Medicine
NRR	National Research Register
OP	Out-patient
PIAA	Physician Insurers Association of America
RCR	Royal College of Radiologists
RDOR	Relative Diagnostic Odds Ratio
RDS	Red Dot System
Regs	Registrars
ROC	Receiver Operating Characteristic curve
SCoR	Society & College of Radiographers
Sens	Sensitivity
SHO	Senior House Officer
SIGLE	System for Information on Grey Literature in Europe
SIGRR	Special Interest Group in Radiographic Reporting
SJH	Saint James' Hospital
Spec	Specificity
TN	True Negative
TP	True Positive
UK	United Kingdom

References

Adrian-Harris D (1997). Workforce Planning in Diagnostic Radiography. In *Current Topics in Radiography*, eds Price R, Patterson A. Saunders.

Aideyan UO, Berbaum K, Smith WL (1995). Influence of Prior Radiological Information on the Interpretation of Radiographic Examinations. *Academic Radiology*, **2**, 205-208.

Altman DG (1991). Clinical Trials. In *Practical Statistics for Medical Research*. Chapman & Hall, London.

Altman DG (1991). Some common problems in medical research. In *Practical statistics for medical research*. Chapman and Hall, London.

Altman DG, Schulz KF, Moher D et al (2001). The Revised CONSORT Statement for Reporting Randomised Trials: Explanation and Elaboration. *Annals Internal Medicine*, **134**, 663-694.

American College of Radiology (1995). *ACR standard for communication: diagnostic radiology*. American College of Radiology, Reston.

Arthur D, Muir J (1909). *A Manual of Practical X-ray Work*. Rebman.

Audit Commission (1995). Improving Your Image: How to Manage Radiology Services More Effectively. Stationary Office, Great Britain.

Bates JA, Conlon RM, Irving HC (1994). An audit of the role of the sonographer in non-obstretic ultrasound. *Clinical Radiology*, **49**, 617-20.

Begg CB (1987). Biases in the assessment of diagnostic tests. *Statistics in Medicine*, **6**, 411-23.

Begg CB, Greenes RA (1983). Assessment of diagnostic tests when verification disease is subject to selection bias. *Biometrics*, **39**, 207-15.

Begg CB, Mazumdar M (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**, 1088-1101.

Beggs I, Davidson JK (1990). Accident and Emergency Reporting in UK Teaching Departments. *Clinical Radiology*, **41**, 264-267.

Berlin JA, Laird NM, Sacks HS, Chalmers TC (1989). A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine*, **8**, 141-151.

Berlin L (1996a). Malpractice Issues in Radiology. The importance of proper radiographic positioning and technique. *American Journal of Roentgenology*, **166**, 769-71.

Berlin L (1996b). Malpractice Issues in Radiology. Perceptual errors. *American Journal of Roentgenology*, **167**, 587-90.

Berlin L (1996c). Malpractice Issues in Radiology. Errors in judgement. *American Journal of Roentgenology*, **166**, 1259-61.

Berlin L (1996d). Malpractice Issues in Radiology. Possessing ordinary knowledge. *American Journal of Roentgenology*, **166**, 1027-9.

Berlin L, Berlin JW (1995). Malpractice and radiologists in Cook County, IL: trends in 20 years of litigation. *American Journal of Roentgenology*, **165**, 781-8.

Berman L, de Lacey G, Craig O (1985b). A survey of accident and emergency reporting: results and implications. *Clinical Radiology*, **36**, 483-484.

Berman L, de Lacey G, Twomey E, Twomey B, Welch T, Eban R (1985a). Reducing errors in accident department; a simple method using radiographers. *British Medical Journal*, **290**, 421-2.

Berman L, de Lacey G, Twomey E, Twomey B, Welch T, Eban R (1986). Reducing errors in the accident department: a simple method using radiographers. *Radiography*, **52**, 143-4.

Bero L, Grilli R, Grimshaw J, Oxman A (Eds.). Effective Practice and Organisation of Care (EPOC) Module of The Cochrane Database of Systematic Reviews (CDSR). In: The Cochrane Library. 1999; Issue 1. Oxford: Updated Software. Updated quaterly.

Brady C (1995). Accountability and the role-development radiographer. *Radiology*, **1**, 127-133.

Brealey S (2001a). Measuring the Effects of Image Interpretation: An Evaluative Framework. *Clinical Radiology*, **56**, 341-347.

Brealey S (2001b). Quality assurance in radiographic reporting: a proposed framework. *Radiography*, **7**, 263-270.

Brealey S, Glenny AM (1999). A framework for radiographers planning to undertake a systematic review. *Radiography*, **5**, 131-146.

Brealey S, King DG, Crowe MTI, Crawshaw I, Ford L, Warnock N, Mannion RAJ, Ethell S (2003). Accident and Emergency and General Practitioner plain radiograph reporting by radiographers and radiologists: A quasi-randomized controlled trial. *British Journal of Radiology*, **76**, 57-61.

Brealey S, Scally AJ (2001). Bias in plain film reading performance studies. *British Journal of Radiology*, **74**, 307-316.

Brealey S, Scally AJ, Thomas NB (2002a). Methodological standards in radiographer plain film reading performance studies. *British Journal of Radiology*, **75**, 107-113.

Brealey S, Scally AJ, Thomas NB (2002b). Presence of bias in radiographer plain film reading performance studies. *Radiography*, **8**, 203 – 210.

Capsticks Solicitors (1994). Hospital and Health Service Law - Focus on risk management. Capsticks (included in the Health Service Journal, Special Report, 20 October 1994), London.

Carr D, Mugglestone M (1997). Visual search strategies of radiographers: evidence for role extension. *European Journal of Radiology*, **7**, 179.

Chalmers TC, Cynthia SF, Reitman D (1990). Minimizing the three stages of publication bias. *Journal of American Medical Association*, **263**, 1392-1395.

Chien PFW, Arnott N,Gordon A, Owen P, Khan KS (2000). How useful is uterine artery doppler flow velocimetry in predicting pre-eclampsia, intrauterine growth retardation and perinatal death? An overview. *British Journal of Obstetrics and Gynaecology*, **107**, 196-208.

Chien PFW, Khan KS (2001). Evaluation of a clinical test. II: Assessment of validity. *British Journal of Obstetrics and Gynaecology*, **108**, 568-572.

Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests: Recommended Methods, updated 6 June 1996. Available at: http://www.cochrane.org/cochrane/sadtdoc1.htm

Cook TD, Campbell DT (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin Company, London.

Craven C, Barber J (1995). Role Extension: Where are we now? *Radiography Today*, **61**, 12-13.

Craven CM, Blanshard KS (1997). Computed tomography (CT) head scans reported by an experienced CT radiographer. *Radiography*, **2**, 105-111.

Cunningham DA (1997). Special Interest Group in Radiographic Reporting. *British Journal of Radiology*, **70**, 873-4.

De Lacey G (1979). Number of casualty attenders referred for X-ray examination. *British Journal of Radiology*, **52**, 52-3.

De Lacey G, Barker A, Harper J, Wignall B (1980). An assessment of the clinical effects of reporting accident and emergency radiographs. *British Journal of Radiology*, **53**, 304-9.

Deeks J (2001a). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, **323**, 157-162.

Deeks J (2001b). Systematic reviews of evaluations of diagnostic and screening tests. In *Systematic Reviews in Health Care: meta-analysis in context*, eds Egger M, Davey-Smith G, Altman DG. British Medical Journal Publishing Group, London.

Deeks J, Khan K, Song F, Popay J, Nixon J, Kleijnen J (2001). Data synthesis (phase 7): Conducting the review (stage II). In *CRD report no.4*. CRD, York.

DeLong ER, DeLong DM, Clarke-Pearson DL (1988). Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, **44**, 837-845.

Department of Health (2000). Meeting the Challenge. A Strategy for the Allied Health Professions. Stationary Office, Great Britain.

Department of Health (2000). The NHS Plan: A Plan for Investment. A Plan for Reform. Stationary Office, Great Britain.

Department of Health (2002). Review Body on Doctors' and Dentists' Remuneration. Thirty-First Report, Stationary Office, Great Britain.

Department of Health (2002). Review Body on Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine. Nineteenth Report on Professions Allied to Medicine, Stationary Office, Great Britain.

Department of Health (2003). *Agenda for Change.* Great Britain: Stationary Office.

Department of Health (2003). *The NHS Knowledge and Skills Framework and Related Development Review.* Great Britain: Stationary Office.

DerSimonian R, Laird N (1986). Meta-analysis in Clinical Trials. *Controlled Clinical Trials*, **7**, 177-188.

Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA (1992). Incorporating variations in the quality of individual randomised trials into meta-analysis. *Journal of Clinical Epidemiology*, **45**, 255-265.

Deville WL, Buntix F, Bouter LM et al (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Medical Research Methodology*, **2**, 9.

Drummond MF, Stoddart GL, Torrance GW (1987). *Methods for the economic evaluation of health care programmes.* Oxford Medical Publications, Oxford.

Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, Matowe L, Needham G, Gilbert F, Bond S (2001). Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet*, **357**, 1406-9.

Egger M, Davey-Smith G, Schneider M, Minder C (1997). Bias in meta-analyses detected by a simple, graphical test. *British Medical Journal*, **315**, 629-634.

Egger M, Dickersin K, Smith GD (2001). Problems and limitations in conducting systematic reviews. In *Systematic Reviews in Health Care: meta-analysis in context*, eds Egger M, Davey-Smith G, Altman DG. British Medical Journal Publishing Group, London.

EuroQol Group (1991). EuroQol: a new facility for the measurement of health related quality of life. *Health Policy*, **16**, 199-208.

Fielding AJ (1990). Improving Accident and Emergency Radiography. *Clinical Radiology*, **41**, 149-151.

Fineberg HV, Bauman R, Sosman M (1977). Computerized cranial tomography: Effect on diagnostic and therapeutic plans. *Journal of the American Medical Association*, **238**, 224-7.

Fryback DG, Thornbury JR (1991). The Efficacy of Diagnostic Imaging. *Medical Decision Making*, **11**, 88-94.

Galbraith RF (1988). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine*, **7**, 889-894.

Garland LH (1949). On the scientific evaluation of diagnostic procedures. *Radiology*; **52**, 309-28.

Gleadhill DNS, Thomson JY, P Simms (1987). Can more efficient use be made of x ray examinations in the accident and emergency department? *British Medical Journal*, **294**, 943-947.

Gore SM (1982). *Statistics in practice.* British Medical Association, London.

Gotzsche P (1989). Multiple publications of reports of drug trials. *European Journal of Clinical Pharmacology*, **36**, 429-432.

Greenland S (1994). A critical look at some popular meta-analytic methods. *American Journal of Epidemiology*, **140**, 290-296.

Grol R (1996). Research and development in quality if care: establishing the research agenda. *Quality in Health Care*, **5**, 235-242.

Haiart DC, Henderson J (1991). A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *British Journal of Clinical Practice*, **45**, 43-5.

Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, et al (1998). Systematic review of endoscopic ultrasound in gastro-oesophageal cancer. *Health Technology Assessment*, **2** (18).

Hasselblad V, Hedges LV (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, **117**, 167-178.

Hughes H, Hughes K, Hamill R (1996). A study to evaluate the introduction of a pattern recognition technique for chest radiographs by radiographers. *Radiography*, **2**, 263-288.

Institute of Medicine (1977). *Policy Statement: Computed Tomographic Scanning.* National Academy of Sciences, Washington DC.

Institute of Medicine (1990). Medicare: A Strategy for Quality Assurance. *Volume 1: Committee to Design a Strategy for Quality Review and Assurance in Medicare*, ed Lohr K. National Academy Press, Washington D.C.

Irwig L, Macaskill P, Glasziou P, Fahey M (1995). Meta-analytic methods for diagnostic accuracy. *Journal of Clinical Epidemiology*, **48**, 119-30.

Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F (1994). Guidelines for Meta-analyses Evaluating Diagnostic Tests. *Annals of Internal Medicine*, **120**, 666-676.

Jaeschke R, Guyatt G, Sackett DL (1994). Users' Guides to the Medical Literature. III. How to Use an Article About a Diagnostic Test. A. Are the Results of the Study Valid? *Journal of American Medical Association*, **271**, 389-391.

James MR, Bracegirdle D, Yates DW (1991). X-ray reporting in accident and emergency departments - an area for improvements in efficiency. *Archives of Emergency Medicine*, **8**, 266-270.

Johnston DA, Brennan PC (2000). Reference dose levels for patients undergoing common diagnostic X-ray examinations in Irish Hospitals. *British Journal of Radiology*, **73**, 396-402.

Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, Hutton J, Smith MA (1997). The identification of bias in studies of the diagnostic performance of imaging modalities. *British Journal of Radiology*, **70**, 1028-1035.

Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M (2001). Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *British Medical Journal*, **322**, 400-5.

Khan KS, Dinnes J, Kleijnen J (2000). Systematic reviews to evaluate diagnostic tests. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **38**, 1-6.

Kido S, Ikezoe J, Takeuchi N, Kondoh H, Tomiyama M, Jokoh T, et al (1993). Interpretation of subtle interstitial lung abnormalities: Conventional versus storage phosphor radiography. *Radiology*, **187**, 527-33.

Kletzenbauer SP (1996). Radiographers' attitudes to skill mix changes. *Radiography*, **2**, 289-300.

Kundel HL (1989). Perception errors in chest radiography. *Seminars in Respiratory Medicine*, **10**, 203-210.

Lafortune M, Breton G, Baudouin JL (1988). The radiological report: what is useful for the referring clinician? *Journal of Canadian Association Radiology*, **39**, 140-143.

Laird NM, Mosteller F (1990). Some statistical methods for combining experimental results. *International Journal of Technology Assessment in Health Care*, **6**, 5-30.

Larkin GV (1983). Occupational Monopoly in Modern Medicine. Tavistock, London.

Last JM (1995). A dictionary of epidemiology, 3rd edn. Oxford University Press, Oxford.

Lijmer JG, Bossuyt PM, Heisterkamp SH (2002). Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine*, **21**, 1525-1537.

Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, Bossuyt, PMM (1999). Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. *Journal of American Medical Association*, **282**, 1061-1066.

Littenberg B, Moses LE (1993). Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. *Medical Decision Making*, **13**, 313-321.

Loughran CF (1994a). Reporting of accident and emergency radiographs by radiographers: a study to determine the effectiveness of a training programme. *British Journal of Radiology*, **67**, 93.

Loughran CF (1994b). Reporting of fracture radiographs by radiographers: the impact of a training programme. *British Journal of Radiology*, **67**, 945-950.

Macaskill P, Glasziou P, Fahey M (1995). Meta-analytic methods for diagnostic test performance. *Journal of Clinical Epidemiology*, **48**, 119-130.

Manning DJ (1998). Evaluation of diagnostic performance in radiography. *Radiography*, **4**, 49-60.

Mannion RAJ, Bewell J, Langan C, et al (1995). A barium enema training programme for radiographers: a pilot study. *Clinical Radiology*, **50**, 715-9.

Maxwell RJ (1984). Quality assessment in health. *British Medical Journal*, **288**, 1470-72.

McAuley L, Ba' Pham, Tugwell P, Moher D (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet*, **356**, 1228-31.

McMaster University Health Science Centre (1981). How to read clinical journals: III. To learn the clinical course and prognosis of disease. *Canadian Medical Association Journal*, **124**, 1156-62.

Meek S, Kendall J, Porter J, Freij R (1998). Can accident and emergency nurse practitioners interpret radiographs? A multicentre study. *Journal of Accident Emergency Medicine*, **15**, 105-7.

Metz CE (1989). Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*, **24**, 234-245.

Midgette AS, Therese A, Stukel A, Littenberg B (1993). A meta-analytic method for summarising diagnostic test performances: Receiver-operating-characteristic – summary point estimates. *Medical Care*, **13**, 253-257.

Moher D, Schulz KF, Altman DG for the CONSORT Group (2001). The CONSORT statment : revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, **357**, 1191-94.

Moodie I (1970). *50 Years of History*. The Society of Radiographers, London.

Moser CA, Kalton G (1979). *Survey Methods in Social Investigation.* Dartmouth, Aldershot.

Moses LE, Shapiro D, Littenberg B (1993). Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine*, **12**, 1293-316.

Mucci B, Lawson S, Athey G, Scarisbrick G (1997). Radiographers as observers in breast screening: Experience with a 'red dot' method. *Breast*, **6**, 183-185.

Mulrow CD (1995). Rationale for systematic reviews. In *Systematic Reviews*, eds Chalmers I, Altman DG. BMJ Publishing Group, London.

Mulrow CD, Linn WD, Gaul MK, Pugh JA (1989). Assessing Quality of a Diagnostic Test Evaluation. *Journal of General Internal Medicine*, **4**, 288-195.

National Health Service and Community Care Act (1990). Stationary Office, Great Britain.

Neisser U (1967). Cognitive psychology. Appleton-Century-Crofts, New York.

Netten A, Curtis L (2002). Unit Costs of Health and Social Care 2002. Personal Social Services Research Unit, Kent.

NHS Centre for Reviews and Dissemination (CRD). Assessing the validity of studies (1996). In: *Undertaking Systematic Reviews of Research on Effectiveness (CRD Report 4)*. The University of York, York.

Ng KH, Rassiah P, Wang HB, Hamboli AS, Muthuvella P, Lee HP (1998). Doses to patients in routine X-ray examinations in Malaysia. *British Journal of Radiology*, **71**, 654-60.

Obuchowski NA, Zepp RC (1996). Simple steps for improving multiple-reader studies in radiology. *American Journal of Roentgenology*, **166**, 517-21.

O'Connor G (1996). Reflection on clinical practice: a means of fostering professional development. *Radiography*, **2**, 53-56.

Overton-Brown P, Anthony D (1998). Towards a partnership in care: nurses' and doctors' interpretation of extremity trauma radiology. *Journal of Advanced Nursing*, **27**, 890-6.

Papageoriou E, Vardalaki E, Hourdakis CJ, Dimitriou P (2001). Estimation of doses received by patients undergoing radiologist examinations in Greece. *Radiation Protection Dosimetry*, **93**, 31-41.

Paterson AM (1995). *Role Development - Towards 2000: A survey of Role Developments in Radiography.* College of Radiographers, London.

Paterson AM, Price R (1996). Current Topics in Radiography Number 2. W.B. Saunders Company Ltd, London.

Pauli R, Hammond S, Cooke J, Ansell J (1996). Radiographers as film observers in screening mammography: An assessment of competence under test and screening conditions. *British Journal of Radiology*, **69**, 10-14.

Physician Insurers Association of America (1997). Practice standards claims survey. PIAA, Maryland.

Piper K, Paterson A, Ryan C (2000). *The Implementation of a Radiographic Reporting Service, for trauma examinations of the skeletal system, in 4 National Health Service Trusts.* Department of Radiography, Canterbury Christ Church University College.

Piper KJ, Paterson AM, Ryan CM (1998). The implementation of a radiographer reporting service for trauma examinations of the skeletal system. *British Journal of Radiology*, **71**, 126.

Pocock SJ (1983). The Justification for Randomized Controlled Trials. In *Clinical Trials - a practical approach.* John Wiley & Sons, Chichester.

Price RC, Le Masurier SB, High L, Miller LR (1999). Changing times: a national survey of extended roles in diagnostic radiography. *British Journal of Radiology*, **72**, 7.

Prime NJ, Paterson AM, Henderson PI (1999). The development of a curriculum – a case study of six courses providing courses in radiographic reporting. *Radiography*, **5**, 63-70.

Radiographers Board of the CPSM (1993). *Statement on Infamous Conduct.* Council for Professions Supplementary to Medicine, London.

Ransohoff DF, Feinstein AR (1978). Problems of spectrum and bias in evaluating the efficacy of the diagnostic tests. N England Journal of Medicine, **199**, 926-30.

Reid MC, Lachs MS, Feinstein AR (1995). Use of Methodological Standards in Diagnostic Test Research. *Journal of American Medical Association*, **274**, 645-651.

Remedios D, Ridley N, Taylor S, de Lacey G (1998). Trauma radiology: extending the Red Dot system. *British Journal of Radiology*, **71**, 60.

Renfrew DL, Franken EA, Berbaum KS, Weigelt FH, Abu-Yousef MM (1992). Errors in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology*, **183**, 145-150.

Renwick IGH, Butt WP, Steele B (1991). How well can radiographers triage X-ray films in the accident and emergency department? *British Medical Journal*, **302**, 568-69.

Richardson G, Maynard A, Cullum N, Kindig D (1998). Skill mix changes: substitution or service development? *Health Policy*, **45**, 119-132.

Rickett AB, Finaly DB, Jagger C (1992). The importance of clinical details when reporting accident and emergency radiographs. *Injury*, **23**, 458-60.

Robinson (1996a). Plain film reporting by radiographers – a feasibility study. *British Journal of Radiology*, **69**, 1171-4.

Robinson PFA, Wilson D, Coral D, Murphy A, Verow P (1999). Variation between experienced observers in the interpretation of accident and emergency radiographs. *British Journal of Radiology*, **72**, 323-330.

Robinson PJA (1996a). Plain film reporting by radiographers - a feasibility study. *British Journal of Radiology*, **69**, 1171-4.

Robinson PJA (1996b). The nature of image reporting. In *Current Topics in Radiography* 2, eds Paterson A, Price R. W B Saunders, London.

Robinson PJA (1997). Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *British Journal of Radiology*, **70**, 1085-1098.

Robinson PJA (1998). Pattern recognition and radiographer reporting. *Radiography*, **4**, 155-157.

Robinson PJA, Culpan G, Wiggins M (1999). Interpretation of selected accident and emergency radiographer examinations by radiographers: a review of 11 000 cases. *British Journal of Radiology*, **72**, 1-6.

Robinson PJA, Fletcher JM (1994). Clinical coding in radiology. *Imaging*, **6**, 133-142.

Rose J, Gallivan S (1991). Plain Film Reporting in the UK. *Clinical Radiology*, **44**, 192-4.

Russell I (1983). The evaluation of computerised tomography: a review of research methods. In *Economical and medical evaluation of health care technologies*, eds Culyer AJ, Horisberger. Springer-Verlag, Berlin.

Russell I, Di Blasa Z, Lambert M, Russell D (1998). In *Systematic reviews and meta-analyses: opportunities and threats*, eds Templeton AA, O'Brien PMS. Evidence-based fertility treatment. RCOG Press, London.

Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991). *Clinical epidemiology: a basic science for clinical medicine*, 3rd edn. Little, Brown and Company, London.

Saxton HM (1992). Should Radiologists Report on Every Film? *Clinical Radiology*, **45**, 1-3.

Scally AJ, Brealey S (2003). Confidence Intervals and Sample Size Calculations for Studies of Film-reading Performance. *Clinical Radiology*, **58**, 238-246.

Secretaries of State for Health, Wales, Northern Ireland and Scotland (1989). Working for Patients. Stationary Office, Great Britain.

Shapiro D (1995). Issues on combining independent estimates of the sensitivity and specificity of a diagnostic test. *Academic Radiology*, **2**, S37-47.

Sierra AE, Bisesi MA, Rosenbaum TL, Potchen EJ (1992). Readability of the Radiologic Report. *Investigative Radiology*, **27**, 236-239.

Simel DL, Feussner JR, Delong ER, Matchar DB (1987). Intermediate, indeterminate, and uninterpretable diagnostic test results. *Medical Decision Making*, **7**, 107-14.

Simel DL, Samsa GP, Matchar DB (1991). Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*; **44**, 763-770.

Smith MJ (1967). *Error and variation in diagnostic radiology*. Charles C. Thomas, Springfield.

Special Interest Group in Radiographic Reporting (2002). *Reporting by Radiographers: A Policy and Practice Guide*. Special Interest Group in Radiographic Reporting.

StataCorp (2001). Stata Statistical Software: Release 7.0. Stata Corporation, Texas.

Sterne JAC, Egger M, Davey-Smith G (2001). Investigating and dealing with publication and other biases in meta-analyses. *British Medical Journal*, **323**, 101-105.

Sterne JAC, Gavaghan D, Egger M (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, **53**, 1119-1129.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000). Meta-analysis of Different Types of Data. In *Methods for Meta-Analysis in Medical Research*. John Wiley & Sons, England.

Swinburne (1971). Pattern recognition for radiographers. *Lancet*, **I**, 589-90.

The College of Radiographers (1993). *Statement of Conduct if the Radiographers Board of the CPSM*. The College of Radiographers, London.

The College of Radiographers (1993). *Code of Practice*. College of Radiographers, London.

The College of Radiographers (1994). *Code of Professional Conduct*. College of Radiographers, London.

The College of Radiographers (1994). *A strategy for research*. College of Radiographers, London.

The College of Radiographers (1995). *X-ray examination requests by nurse practitioners and radiographic reporting by radiographers*. College of Radiographers, London.

The College of Radiographers (1996). *Role Development in Radiography*. College of Radiographers, London.

The College of Radiographers (1997). *Reporting by radiographers: A vision paper*. College of Radiographers, London.

The ISRRT (1999). Survey on Conditions for the Education of Radiographers in Europe.

The Royal College of Radiologists & College of Radiographers (1998). *Inter-Professional Roles and Responsibilities in a Radiology Service*. The Royal College of Radiologists & College of Radiographers, London.

The Royal College of Radiologists (1993a). *Medical staffing and workload in clinical radiology in the United Kingdom NHS*. Royal College of Radiologists, London.

The Royal College of Radiologists (1993b). *Staffing and Standards in Departments of Clinical Oncology and Clinical Radiology*. Royal College of Radiologists, London.

The Royal College of Radiologists (1993c). *Medico Legal Aspects of Delegation*. Royal College of Radiologists, London.

The Royal College of Radiologists (1994). Radiographer role development in clinical radiology; Memorandum. London.

The Royal College of Radiologists (1994). *Röntgen's Progress. A Discussion Paper on the Future of Clinical Radiology in the UK*. Royal College of Radiologists, London.

The Royal College of Radiologists (1995a). *Statement on reporting in Departments of Clinical Radiology*. Royal College of Radiologists, London.

The Royal College of Radiologists (1995b). *Clinical Radiology Quality Specification for Purchasers*. London: The Royal College of Radiologists.

The Royal College of Radiologists (1996). *Diagnostic Imaging and the Primary Care Sector*. London: The Royal College of Radiologists.

The Royal College of Radiologists (1999). *Skill Mix in Clinical Radiology*. London: The Royal College of Radiologists.

The Royal College of Radiologists (2000). *Two National Audits of Radiographic Reporting Services*. London: The Royal College of Radiologists.

The Royal College of Radiologists (2002). *Clinical Radiology: A Workforce in Crisis*. London: The Royal College of Radiologists.

The Royal College of Radiologists Working Party (1991). A multicentre audit of hospital referral for radiological investigation in England and Wales. British Medical Journal, **303**, 809-812.

The Society and College of Radiographers (1994). X-Ray Examinations by Nurse Practitioners and Radiographer reporting by Radiographers. The Society and College of Radiographers, London.

Thomas HG, Mason AC, Smith RM, Fergusson CM (1992). Value of radiographic audit in an accident service department. *Injury*, **23**, 47-50.

Thompson SG (1994). Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*, **309**, 1351-1355.

Tudor GR, Finlay D, Taub N (1997). An assessment of inter-observer agreement and accuracy when reporting plain radiographs. *Clinical Radiology*, **52**, 235-8.

van der Schouw YT, Verbeek ALM, Ruijs SHJ (1995). Guidelines for the Assessment of New Diagnostic Tests. *Investigative Radiology*, **30**, 334-340.

Wardrope J, Chennels PM (1985). Should all casualty radiographs be reviewed? *British Medical Journal*, **290**, 1638-40.

Webster J, Gallagher D (1998). Radiography reporting in a community hospital. *British Journal of Radiology*, **70**, 42.

White K, Berbaum KS, Smith WL (1994). Role of previous radiographs and reports in the interpretation of current radiographs. *Investigative Radiology*, **29**, 263.

Whitehead A, Whitehead J (1991). A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine*, **10**, 1665-1677.

Wittenberg J, Fineberg HV, Ferraci JT et al (1980). Clinical efficacy of computed body tomography. *American Journal of Roentgenology*, **134**, 111-20.