This is a repository copy of *Imaging tests for the detection of osteomyelitis:a systematic review*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/153018/

Version: Published Version

Article:

Llewellyn, Alexis orcid.org/0000-0003-4569-5136, Jones-Diette, Julie orcid.org/0000-0003-1769-8612, Kraft, Jeannette et al. (3 more authors) (2019) Imaging tests for the detection of osteomyelitis:a systematic review. Health technology assessment. pp. 1-128. ISSN 2046-4924

https://doi.org/10.3310/hta23610

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

NIHR | National Institute
for Health Research

Journals Library

Health Technology Assessment
Volume 23 • Issue 61 • October 2019
ISSN 1366-5278

Imaging tests for the detection of osteomyelitis: a systematic review

Alexis Llewellyn, Julie Jones-Diette, Jeannette Kraft, Colin Holton,
Melissa Harden and Mark Simmonds

DOI 10.3310/hta23610

Imaging tests for the detection of osteomyelitis: a systematic review

Alexis Llewellyn,[1] Julie Jones-Diette,[1]
Jeannette Kraft,[2] Colin Holton,[2] Melissa Harden[1]
and Mark Simmonds[1]*

[1]Centre for Reviews and Dissemination, University of York, York, UK
[2]Leeds Teaching Hospitals NHS Trust, Leeds, UK

*Corresponding author

Declared competing interests of authors: none

Published October 2019
DOI: 10.3310/hta23610

This report should be referenced as follows:

Llewellyn A, Jones-Diette J, Kraft J, Holton C, Harden M, Simmonds M. Imaging tests for the detection of osteomyelitis: a systematic review. *Health Technol Assess* 2019;**23**(61).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents*®/ Clinical Medicine.

Health Technology Assessment HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 16/103/03. The contractual start date was in July 2017. The draft report began editorial review in October 2018 and was accepted for publication in March 2019. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Senior Clinical Researcher, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Imaging tests for the detection of osteomyelitis: a systematic review

Alexis Llewellyn,[1] Julie Jones-Diette,[1] Jeannette Kraft,[2] Colin Holton,[2] Melissa Harden[1] and Mark Simmonds[1]*

[1]Centre for Reviews and Dissemination, University of York, York, UK
[2]Leeds Teaching Hospitals NHS Trust, Leeds, UK

*Corresponding author mark.simmonds@york.ac.uk

Background: Osteomyelitis is an infection of the bone. Medical imaging tests, such as radiography, ultrasound, magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT) and positron emission tomography (PET), are often used to diagnose osteomyelitis.

Objectives: To systematically review the evidence on the diagnostic accuracy, inter-rater reliability and implementation of imaging tests to diagnose osteomyelitis.

Data sources: We conducted a systematic review of imaging tests to diagnose osteomyelitis. We searched MEDLINE and other databases from inception to July 2018.

Review methods: Risk of bias was assessed with QUADAS-2 [quality assessment of diagnostic accuracy studies (version 2)]. Diagnostic accuracy was assessed using bivariate regression models. Imaging tests were compared. Subgroup analyses were performed based on the location and nature of the suspected osteomyelitis. Studies of children, inter-rater reliability and implementation outcomes were synthesised narratively.

Results: Eighty-one studies were included (diagnostic accuracy: 77 studies; inter-rater reliability: 11 studies; implementation: one study; some studies were included in two reviews). One-quarter of diagnostic accuracy studies were rated as being at a high risk of bias. In adults, MRI had high diagnostic accuracy [95.6% sensitivity, 95% confidence interval (CI) 92.4% to 97.5%; 80.7% specificity, 95% CI 70.8% to 87.8%]. PET also had high accuracy (85.1% sensitivity, 95% CI 71.5% to 92.9%; 92.8% specificity, 95% CI 83.0% to 97.1%), as did SPECT (95.1% sensitivity, 95% CI 87.8% to 98.1%; 82.0% specificity, 95% CI 61.5% to 92.8%). There was similar diagnostic performance with MRI, PET and SPECT. Scintigraphy (83.6% sensitivity, 95% CI 71.8% to 91.1%; 70.6% specificity, 57.7% to 80.8%), computed tomography (69.7% sensitivity, 95% CI 40.1% to 88.7%; 90.2% specificity, 95% CI 57.6% to 98.4%) and radiography (70.4% sensitivity, 95% CI 61.6% to 77.8%; 81.5% specificity, 95% CI 69.6% to 89.5%) all had generally inferior diagnostic accuracy. Technetium-99m hexamethylpropyleneamine oxime white blood cell scintigraphy (87.3% sensitivity, 95% CI 75.1% to 94.0%; 94.7% specificity, 95% CI 84.9% to 98.3%) had higher diagnostic accuracy, similar to that of PET or MRI. There was no evidence that diagnostic accuracy varied by scan location or cause of osteomyelitis, although data on many scan locations were limited. Diagnostic accuracy in diabetic foot patients was similar to the overall results. Only three studies in children were identified; results were too limited to draw any conclusions. Eleven studies evaluated inter-rater reliability. MRI had acceptable inter-rater reliability. We found only one study on test implementation and no evidence on patient preferences or cost-effectiveness of imaging tests for osteomyelitis.

Limitations: Most studies included < 50 participants and were poorly reported. There was limited evidence for children, ultrasonography and on clinical factors other than diagnostic accuracy.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Conclusions: Osteomyelitis is reliably diagnosed by MRI, PET and SPECT. No clear reason to prefer one test over the other in terms of diagnostic accuracy was identified. The wider availability of MRI machines, and the fact that MRI does not expose patients to harmful ionising radiation, may mean that MRI is preferable in most cases. Diagnostic accuracy does not appear to vary with the potential cause of osteomyelitis or with the body part scanned. Considerable uncertainty remains over the diagnostic accuracy of imaging tests in children. Studies of diagnostic accuracy in children, particularly using MRI and ultrasound, are needed.

Study registration: This study is registered as PROSPERO CRD42017068511.

Funding: This project was funded by the National Institute for Health Research Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 23, No. 61. See the NIHR Journals Library website for further project information.

Contents

List of tables ix

List of figures xi

Glossary xiii

List of abbreviations xv

Plain English summary xvii

Scientific summary xix

Chapter 1 Background **1**
Osteomyelitis 1
Diagnostic imaging for osteomyelitis 1
Current diagnostic and treatment practice 2
Pathway to diagnosis in the NHS 2
Existing review evidence 3

Chapter 2 Aims and objectives **5**

Chapter 3 Methods **7**
Literature searches 7
Study selection 7
  Participants 7
  Index tests 7
  Reference standards 8
  Outcomes 8
  Study designs 8
Data extraction 8
Quality assessment 9
Synthesis 9
  Diagnostic meta-analysis 9
  Inter-rater reliability and implementation review 11
  Deviations from the protocol 11
Role of patient and clinical advisors 11

Chapter 4 Results **13**
Quantity and quality of research available 13
  Included studies 13
  Excluded studies 14
Assessment of diagnostic accuracy 14
  Characteristics of included studies 14
  Critical appraisal 19
Synthesis of diagnostic accuracy in adults 22
  Comparisons between tests 24
  Synthesis of specific imaging tests 26
  Other factors and subgroups 32

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Synthesis of diagnostic accuracy in people with diabetic foot ulcers 35
Comparisons of tests 36
Synthesis of diagnostic accuracy for osteomyelitis not related to diabetes 37
Comparisons of tests 39
Studies not included in the quantitative synthesis 39
Synthesis of studies in children 40
Review of inter-rater reliability 42
Characteristics of included studies 42
Critical appraisal 42
Results of inter-rater reliability studies 44
Review of implementation 44
Summary of previous systematic reviews 46
Clinical effectiveness summary and conclusions 46
Summary of included studies 46
General conclusions from synthesis 51
Diagnostic accuracy by cause or nature of osteomyelitis 51
Studies of children, inter-rater reliability and implementation 52

Chapter 5 Discussion **53**
Statement of principal findings 53
Diagnostic accuracy in adults 53
Diagnostic tests in children 54
Inter-rater reliability and implementation 54
Strengths and limitations of the assessment 54
Strengths 54
Limitations 55
Uncertainties 56
Considerations from patient representatives 56

Chapter 6 Conclusions **57**
Implications for health care 57
Suggested research priorities 58

Acknowledgements **59**

References **61**

Appendix 1 Literature searches **73**

Appendix 2 Results of critical appraisal of diagnostic accuracy studies (QUADAS-2) **91**

Appendix 3 Participant selection criteria in diagnostic accuracy studies **105**

Appendix 4 Results of diagnostic meta-analyses **111**

List of tables

TABLE 1 Study design and participant characteristics 15

TABLE 2 Diagnostic accuracy studies according to test and body part targeted 19

TABLE 3 Summary of critical appraisal of diagnostic accuracy studies 19

TABLE 4 Results of univariate meta-analyses 26

TABLE 5 Summary of regression model to compare imaging tests 27

TABLE 6 Summary of diagnostic accuracy of ultrasound studies 30

TABLE 7 Comparison of tests in people with diabetic foot ulcers 36

TABLE 8 Comparison of tests in people without diabetes 39

TABLE 9 Studies not included in the quantitative synthesis 40

TABLE 10 Diagnostic accuracy results of studies in children 41

TABLE 11 Characteristics of inter-rater reliability studies 43

TABLE 12 Critical appraisal of inter-rater reliability studies 44

TABLE 13 Results of inter-rater reliability studies 45

TABLE 14 Summary of previous systematic reviews 47

TABLE 15 The QUADAS-2 signalling questions for risk-of-bias assessment 91

TABLE 16 The QUADAS-2 risk-of-bias decisions 92

TABLE 17 The QUADAS-2 applicability decisions 102

TABLE 18 Participant selection criteria in diagnostic accuracy studies 105

TABLE 19 Summary of diagnostic accuracy studies 111

TABLE 20 Results from logistic regression model to compare tests 115

TABLE 21 Summary of studies of scintigraphy 115

TABLE 22 Results from logistic regression model to compare tests 118

TABLE 23 Univariate meta-analyses of studies of people with diabetic foot ulcers 122

TABLE 24 Results of logistic regression to compare tests – diabetic foot ulcers 123

TABLE 25 Univariate meta-analysis – patients without diabetes 126

TABLE 26 Results of logistic regression to compare tests – patients without diabetes 127

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

List of figures

FIGURE 1 A PRISMA flow diagram showing study selection 13

FIGURE 2 Summary ROC plot of diagnostic accuracy in adults 22

FIGURE 3 Bivariate meta-analysis of sensitivity and specificity: all adult studies 23

FIGURE 4 Summary HSROC curves: all adult studies 23

FIGURE 5 Forest plots of sensitivity and specificity for MRI 25

FIGURE 6 Comparing DORs within studies 27

FIGURE 7 Summary ROC plot for scintigraphy studies 28

FIGURE 8 Diagnostic accuracy of scintigraphy according to type of test performed 29

FIGURE 9 Diagnostic accuracy of SPECT studies according to type of SPECT used 30

FIGURE 10 Diagnostic accuracy results for combination tests 31

FIGURE 11 Diagnostic odds ratios for combination tests 31

FIGURE 12 Sensitivity and specificity according to whether or not any imaging test was used before the main test 33

FIGURE 13 Sensitivity and specificity according to whether the study was of AOM or COM 34

FIGURE 14 Bivariate meta-analysis according to acute or chronic status 34

FIGURE 15 Sensitivity and specificity in studies of patients with diabetic foot ulcers 35

FIGURE 16 Bivariate analysis of sensitivity and specificity in people with diabetic foot ulcers 36

FIGURE 17 Sensitivity and specificity according to scan location 37

FIGURE 18 Sensitivity and specificity according to likely cause of osteomyelitis 38

FIGURE 19 Bivariate meta-analysis in studies of people without diabetes 39

FIGURE 20 Summary of risk of bias and applicability assessment 91

FIGURE 21 Summary of PPV and NPV in adults 113

FIGURE 22 Summary of PPV and NPV from bivariate analysis 113

FIGURE 23 Comparisons of DORs within studies 114

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

FIGURE 24 Association between specificity and incidence of osteomyelitis in MRI studies 117

FIGURE 25 Sensitivity and specificity for PET and PET-CT studies 117

FIGURE 26 Sensitivity and specificity according to test and reference standard used 118

FIGURE 27 Sensitivity and specificity according to risk of bias owing to patient selection 119

FIGURE 28 Sensitivity and specificity according to risk of bias owing to index test 119

FIGURE 29 Sensitivity and specificity according to presence of indwelling metalwork 120

FIGURE 30 The PPV and NPV in studies of people with diabetic foot ulcers 120

FIGURE 31 The HSROC curves in studies of people with diabetic foot ulcers 121

FIGURE 32 Summary PPV and NPV in studies of people with diabetic foot ulcers 121

FIGURE 33 Meta-analysis of scintigraphy studies in people with diabetic foot ulcers 122

FIGURE 34 Difference in DOR between studies – patients with diabetic foot ulcers 123

FIGURE 35 The PPV and NPV by test location 124

FIGURE 36 The PPV and NPV by cause of osteomyelitis 124

FIGURE 37 Summary sensitivity and specificity by test location 125

FIGURE 38 Summary sensitivity and specificity by cause of osteomyelitis 125

FIGURE 39 Comparisons of DORs across tests – patients without diabetes 127

Glossary

Diagnostic odds ratio Ratio of odds of sensitivity to odds of specificity (summarises diagnostic accuracy).

Diphosphono-1,2-propanodicarboxylic acid A substance used in scintigraphy.

False negative Incorrect negative test result – an affected individual with a negative test result.

False positive Incorrect positive test result – an unaffected individual with a positive test result.

Fludeoxyglucose A radiopharmaceutical used in positron emission tomography.

Hexamethylpropyleneamine oxime A substance used in scintigraphy.

Histopathology The microscopic study of tissue samples to enable the diagnosis of osteomyelitis.

(Hydroxy)methylene diphosphonate A substance used in scintigraphy.

Index test The test for which diagnostic accuracy is being evaluated.

Indium-111 A radioisotope used in scintigraphy.

Meta-analysis Statistical (or quantitative) techniques used to combine the results of two or more studies and obtain a combined estimate of effect.

Microbiology A laboratory method used to diagnose osteomyelitis; a microbiological culture involves multiplying microbial organisms by letting them reproduce under controlled laboratory conditions.

Negative predictive value Proportion of people who test negative (for osteomyelitis) who do not have the condition.

Positive predictive value Proportion of people who test positive who have the condition.

Positive rate Proportion of people with a positive index test result (i.e. who might be diagnosed with osteomyelitis).

Sensitivity Proportion of people with the condition (osteomyelitis) who are correctly diagnosed.

Specificity Proportion of people without the condition who are correctly diagnosed.

Technetium-99m A radioisotope used in scintigraphy and single-photon emission computed tomography.

True negative A correct negative test result – an unaffected individual with a negative test result.

True positive A correct positive test result – an affected individual with a positive test result.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

List of abbreviations

^{18}F-FDG	fludeoxyglucose
3D	three-dimensional
^{99m}Tc	technetium-99m
AGA	antigranulocyte antibody
AOM	acute osteomyelitis
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CI	confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health
COM	chronic osteomyelitis
CRD	Centre for Reviews and Dissemination
CT	computed tomography
DARE	Database of Abstracts of Reviews of Effects
DOR	diagnostic odds ratio
DPD	diphosphono-1, 2-propanodicarboxylic acid
GP	general practitioner
(H)MDP	(hydroxy)methylene diphosphonate
HMPAO	hexamethylpropyleneamine oxime
HSROC	hierarchical summary receiver operating characteristic
HTA	Health Technology Assessment
MDP	methylene diphosphonate
MeSH	medical subject heading
MRI	magnetic resonance imaging
NICE	National Institute for Health and Care Excellence
NPV	negative predictive value
PET	positron emission tomography
PPV	positive predictive value
PR	positive rate
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2	quality assessment of diagnostic accuracy studies (version 2)
ROC	receiver operating characteristic
SPECT	single-photon emission computed tomography
WBC	white blood cell

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Plain English summary

Osteomyelitis is an infection of the bone and is treated with antibiotics. Left untreated, it can cause permanent damage and can lead to amputation.

The best method to diagnose osteomyelitis is to take a bone sample (bone biopsy) but this is invasive and painful. Imaging may help target the best locations for biopsies or remove the need for a biopsy entirely. Several methods are available, including radiography, ultrasound, magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT) and positron emission tomography (PET).

This project systematically reviewed the relevant literature to determine which tests are the most accurate and relevant for clinical practice. All types of patients and all types of osteomyelitis were reviewed. Studies were pooled using statistical methods (meta-analyses) to estimate the overall accuracy of the imaging tests.

The review identified 81 studies and concluded that MRI, PET and SPECT all had similar accuracy, correctly identifying over 85% of people who did have osteomyelitis and over 80% of people who did not have osteomyelitis. Radiography and computed tomography were less accurate. Modern forms of scintigraphy have accuracy similar to PET or MRI.

There was no evidence that the accuracy of the imaging tests was different depending on the cause of osteomyelitis or which body part was affected. In particular, diagnostic accuracy in people with diabetic foot ulcers was similar to other types of osteomyelitis in adults. There was not enough evidence about which tests are most accurate in children, so further studies in children are needed.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Scientific summary

Background

Osteomyelitis is an infection of the bone and bone marrow that may result in bone infarction and loss of limb or joint function and, in extreme cases, may necessitate amputation of the affected limb. In children, osteomyelitis may also inhibit limb growth. Osteomyelitis is common in people with vascular deficiency, such as adults with diabetes mellitus.

Patients usually present with a range of symptoms including swelling, joint pain and fever. These symptoms are often not specific to osteomyelitis, leading to delays in correct diagnosis. Blood tests are used initially to assess inflammatory markers; when these tests show evidence of possible infection, patients are referred for further diagnostic testing. The most accurate diagnostic tool is a bone biopsy or aspiration of a pus collection from the bone or tissue surrounding the bone, with a histological and/or microbiological assessment of the sample to identify the organism causing the infection. The primary treatment for osteomyelitis is a course of antibiotics, but surgery may also be used.

Diagnostic imaging for osteomyelitis

A range of diagnostic imaging methods are available, including radiography, magnetic resonance imaging (MRI) scans, computed tomography (CT) scans, scintigraphy, positron emission tomography (PET) scans, single-photon emission computed tomography (SPECT) and ultrasound.

Little formal guidance [such as guidelines produced by the National Institute for Health and Care Excellence (NICE)] exists for which imaging techniques to use to diagnose osteomyelitis. The only current NICE guidance is for the treatment of diabetic foot ulcers. In those patients, radiography is recommended, followed by MRI if osteomyelitis is suspected, but not confirmed, by radiography.

Objectives

The key objectives were to:

- perform a systematic review of all studies reporting the diagnostic accuracy of any relevant imaging test, or combination of tests used to detect osteomyelitis
- perform diagnostic meta-analyses of identified studies to formally assess their diagnostic accuracy
- investigate diagnostic accuracy across the range of different types of osteomyelitis and types of patient
- compare the diagnostic accuracy of diagnostic tests both statistically and pragmatically by systematically reviewing inter-rater reliability and also the broader issues around implementation of imaging tests, such as availability of machinery, radiation exposure and acceptability to patients
- provide useful guidance as to which imaging tests should be preferred, according to type of disease and patient, in the UK.

Methods

A systematic review of the clinical effectiveness was performed following the general principles recommended in the University of York Centre for Reviews and Dissemination's (CRD's) guidance and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. The protocol details have been registered on PROSPERO (number CRD42017068511).

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Literature searches

Comprehensive searches for published and unpublished literature were carried out during August 2017 and updated in July 2018. Databases searched included MEDLINE and EMBASE.

Study selection

Titles and abstracts and the full texts of studies were independently assessed for inclusion by two reviewers using the inclusion criteria outlined in the following sections.

Participants

Participants included any patients with suspected osteomyelitis (based on symptoms, surgical samples or blood tests). No restrictions were made for age or disease aetiology.

Index tests

Index texts included any diagnostic imaging technique that could potentially identify osteomyelitis, including radiography, MRI, CT, PET, scintigraphy, SPECT and ultrasound.

Reference standards

Histopathology or microbiology based on bone biopsy or pus aspiration, and surgery, were the reference standards. As biopsies are invasive, clinical follow-up of at least 6 months with no signs or symptoms of osteomyelitis was also accepted as confirmation of the absence of osteomyelitis.

Outcomes

Studies reporting diagnostic accuracy of imaging tests compared with a reference standard expressed in terms of sensitivity (percentage with osteomyelitis with a positive diagnostic test result) and specificity (percentage without osteomyelitis with a negative test result) were included.

Studies reporting inter-rater reliability data or other data on test interpretation were included. General implementation outcomes considered were cost-effectiveness (of relevance to the UK), availability of tests, radiation exposure and experience of patients and clinicians.

Study designs

Any study that considered an imaging test or tests for osteomyelitis and which reported data on any of the specified outcomes was included. Only studies explicitly considering testing for osteomyelitis were included.

Data extraction

Study and patient characteristics were extracted by at least one reviewer and checked by a second reviewer. The numbers of true-positive, true-negative, false-positive and false-negative test results were extracted where possible.

Inter-rater reliability estimates were extracted from the papers and tabulated. For implementation studies, relevant results (e.g. from surveys of clinicians) were extracted and summarised narratively.

Quality assessment

The quality of the included diagnostic accuracy studies was assessed using the QUADAS-2 [quality assessment of diagnostic accuracy studies (version 2)] tool. Critical appraisal was performed by one reviewer and independently checked by another.

Synthesis

Data were synthesised in meta-analyses across studies using logistic regression modelling. Random-effects models were used to account for potential heterogeneity in diagnostic accuracy across studies. Results were presented as summary sensitivity and specificity estimates, with 95% confidence intervals (CIs), and as summary HSROC (hierarchical summary receiver operating characteristic) curves. Analyses were performed separately for adults and children.

When the studies were deemed too diverse for meta-analysis to be suitable, or where only one or two studies were available, the reported diagnostic accuracy from each available study was presented in tables and on ROC (receiver operating characteristic) plots.

Separate meta-analyses were conducted for each diagnostic imaging test and, when sufficient data were available, in subcategories of patients including:

- patients with diabetic foot ulcers
- cause of osteomyelitis
- anatomical site.

When studies report diagnostic accuracy data for two or more imaging tests, these tests were compared by extending the bivariate logistic regression models to include all imaging tests in one model.

Inter-rater reliability results and qualitative data on implementation were reported narratively and tabulated. Areas where few or no data have been published were also identified.

Results

Diagnostic accuracy

The review of diagnostic accuracy included 77 studies. The sample size of the studies ranged from 7 to 339, but most (80%) included fewer than 50 participants. Nearly one-quarter of the studies were considered as being at a high risk of bias, although poor reporting meant that there was significant uncertainty about the quality of most studies. Most of the evidence focused on the diagnostic accuracy of MRI, scintigraphy and radiography for the diagnosis of diabetic foot osteomyelitis. Few studies specifically focused on the axial skeleton, the pelvis/hip/knee and long bones.

The overall meta-analysis of diagnostic accuracy in adults found that MRI can detect osteomyelitis with high accuracy (95.6% sensitivity, 95% CI 92.4% to 97.5%; 80.7% specificity, 95% CI 70.8% to 87.8%). PET also had high diagnostic accuracy (85.1% sensitivity, 95% CI 71.5% to 92.9%; 92.8% specificity, 95% CI 83.0% to 97.1%), as did SPECT (95.1% sensitivity, 95% CI 87.8% to 98.1%; 82.0% specificity, 95% CI 61.5% to 92.8%). There were similar diagnostic odds ratios and summary HSROC curves for MRI, PET and SPECT, suggesting that the three imaging tests have similar diagnostic performance.

Scintigraphy (83.6% sensitivity, 95% CI 71.8% to 91.1%; 70.6% specificity, 95% CI 57.7% to 80.8%), CT (69.7% sensitivity, 95% CI 40.1% to 88.7%; 90.2% specificity, 95% CI 57.6 to 98.4) and radiography (70.4% sensitivity, 95% CI 61.6% to 77.8%; 81.5% specificity, 95% CI 69.6% to 89.5%) all had generally inferior diagnostic accuracy when compared with MRI, PET or SPECT. The most up-to-date forms of scintigraphy, such as ^{99m}Tc HMPAO WBC (technetium-99m hexamethylpropyleneamine oxime white blood cell) scintigraphy (87.3% sensitivity, 95% CI 75.1% to 94.0%; 94.7% specificity 95% CI 84.9% to 98.3%), had high diagnostic accuracy, similar to that of PET or MRI. There were insufficient studies of ultrasound to assess its diagnostic accuracy.

Key participant subgroups

The main patient subgroup was patients with diabetic foot ulcers, representing nearly half of all studies. The results of the meta-analyses for these patients were similar to those from the main meta-analysis, although there were too few studies of SPECT or CT to reliably assess diagnostic accuracy.

Studies of patients without diabetes were divided according to scan location and by potential cause of osteomyelitis. Data within each category were generally limited, but there was no evidence that diagnostic accuracy varied by scan location or cause, or that results differed substantially from the main analysis.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Diagnostic tests in children

The evidence for the accuracy of ultrasound and MRI in children was mixed and limited overall. Ultrasonography had moderate sensitivity and specificity in children with suspected acute haematogenous osteomyelitis but had perfect sensitivity and specificity in a small study of children with negative or equivocal initial radiographic findings. MRI had good sensitivity and specificity in children with suspected acute haematogenous osteomyelitis in one study, but preoperative MRI had poor sensitivity and near perfect specificity in another study of patients with septic hip.

Inter-rater reliability and implementation

Eleven studies evaluated the inter-rater reliability of at least one imaging test, and one study provided data on clinician opinions on imaging tests for osteomyelitis.

Magnetic resonance imaging appeared to have acceptable inter-rater reliability. There was some evidence suggesting that PET and scintigraphy showed near perfect inter-rater reliability, although this is limited to two small studies. We found no evidence on patient preferences and cost-effectiveness of imaging tests for osteomyelitis.

Only one study on the implementation of diagnostic test imaging for osteomyelitis was included. A Dutch survey of clinicians found that preferred imaging strategies for diagnosing post-traumatic osteomyelitis depended on specialty and availability of machinery. Most responders were not aware of local hospital protocols for diagnosing osteomyelitis.

Discussion

Strengths and limitations of the analyses

This systematic review was the largest and most comprehensive review of the diagnosis of osteomyelitis to date, and the first to comprehensively compare all relevant imaging tests across all types of patient.

There were few studies identified that included children. This may be because studies of children do not discuss osteomyelitis directly, but instead mostly in the context of other conditions, such as septic arthritis. Hence, it is possible that some possibly relevant studies were missed.

Some imaging tests were reported, particularly ultrasound and CT scans. Some aspects of studies were inconsistently reported, such as varying descriptions of the cause of osteomyelitis, non-reporting of whether osteomyelitis was acute or chronic and lack of clarity on whether or not radiography (or other tests) had been used prior to the main test. This made assessment in these subgroups difficult.

We identified very few data beyond those on diagnostic accuracy, with few studies discussing broader implementation issues such as access to machinery, costs or radiation exposure.

Uncertainties

The main uncertainties remaining following this review arise largely because of limitations in the identified studies.

The diagnostic accuracy of imaging tests in children remains highly uncertain because of the very limited nature of the evidence. We could reach no firm conclusions on the diagnostic accuracy of any imaging test in children.

The diagnostic accuracy of ultrasound is currently unknown as the only two studies in adults had conflicting results.

Although we found no evidence that diagnostic accuracy varied across subgroups of patients, limited or inconsistent reporting of some characteristics, such as acute versus chronic osteomyelitis, or the cause of osteomyelitis, means than differences between tests or between subgroups cannot be ruled out.

Generalisability of the findings

The apparent consistency of diagnostic accuracy across the various types of patient and causes of osteomyelitis suggests that the diagnostic accuracy findings are likely to be generalisable to any population being tested for osteomyelitis. The similarity in diagnostic accuracy across MRI, PET and SPECT scans suggests that these tests should have similar accuracy in most clinical circumstances.

The review found considerable variation in the specificity of MRI across studies. This may mean that the observed specificity of MRI in any given setting may differ from the summary estimates calculated in this review, depending on how MRI is implemented.

The limited evidence on diagnosis of osteomyelitis in children means that results may not be generalisable beyond the populations in the included studies.

Conclusions

Implications for service provision

Magnetic resonance imaging, PET and SPECT all have broadly similar and high accuracy when diagnosing osteomyelitis. All are likely to be suitable imaging tests for diagnosing osteomyelitis. No clear reason to prefer one test over the others in terms of diagnostic accuracy was identified. The wider availability of MRI machines, and the fact that MRI does not expose patients to harmful ionising radiation, may mean that MRI is preferable in most cases, unless it is unsuitable for a particular patient. A PET or SPECT scan may be required if a MRI scan is inconclusive.

Positron emission tomography had poorer sensitivity but higher specificity than MRI, with more consistent results across studies. This may make PET better suited to situations where avoiding false-positive diagnoses is important, for example when the test would be followed by surgery or other invasive procedures.

There is no evidence to suggest that the diagnostic accuracy varies with the potential cause of osteomyelitis or with the body part scanned, although data on patients other than those with diabetic foot ulcers were limited. The review identified very limited data on diagnosing osteomyelitis in children, so considerable uncertainty remains over the diagnostic accuracy of imaging tests in children. Clinicians should be aware of this limitation in the evidence base.

Suggested research priorities

The most urgent research priority is to perform diagnostic accuracy studies of imaging tests in children. Large diagnostic accuracy studies are needed, which must be of high quality, with proper blinding of test assessors and consecutive recruitment of patients. The priority tests should be MRI and ultrasound, ideally comparing the two tests in the same children.

Ultrasound has not been widely assessed in adults. Current results suggest that ultrasound on its own may not be sufficiently accurate to diagnose osteomyelitis, but further accuracy studies are needed to resolve the uncertainty. It may be more appropriate to investigate the diagnostic accuracy of ultrasound as a precursor to MRI or other tests (e.g. as a replacement for radiography).

Given the similarities in diagnostic accuracy of MRI, PET and SPECT, suitable investigation of patient and clinician experience and opinion of these tests, through surveys or focus groups, would be useful to identify practical reasons for the choice of test. Similarly, a formal economic evaluation of these tests, accounting for test cost, availability and risk of radiation exposure, would help to clarify the choice between these tests.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Study registration

This study is registered as PROSPERO CRD42017068511.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Background

Osteomyelitis

Osteomyelitis is an infection of the bone and bone marrow.[1,2] Left untreated, it may result in bone infarction and loss of limb or joint function and, in extreme cases may require amputation of the affected limb. If the infection spreads, it may lead to potentially fatal septicaemia.[3] In children, osteomyelitis may also inhibit limb growth, requiring extensive orthopaedic intervention in later childhood. *Staphylococcus aureus* is the most common organism causing osteomyelitis, but other common organisms such as *Streptococcus* spp. or *Escherichia coli* may also be responsible in some cases. Bone infections occur most commonly in people aged < 20 years or > 50 years. It accounts for around 1% of all childhood hospital admissions. The incidence of osteomyelitis has increased over recent decades, notably in children and in patients > 60 years of age. This growing incidence has been associated with increased prevalence of meticillin-resistant *S. aureus* (MRSA) in children and an increase in diabetes mellitus-related infections in adults.[4]

Osteomyelitis may be acute, subacute or chronic and is divided between haematogenous osteomyelitis, in which infection transfers from a remote location in the body via the bloodstream, and contiguous osteomyelitis, in which infected material comes into direct contact with the bone.[5] The haematogenous type is more common in children whereas the contiguous type is more common in adults, usually as a result of trauma or surgery.[6] Osteomyelitis is also common in people with vascular deficiency, such as adults with diabetes, as a complication of diabetic foot ulcers.[7] Osteomyelitis may lead to infection of the adjacent joint (septic arthritis) or occur secondary to septic arthritis by contiguous spread.

Patients usually present with a range of symptoms including swelling, joint pain and fever. These symptoms are often not specific to osteomyelitis, leading to delays in correct diagnosis. Blood tests are used initially to assess inflammatory markers indicative of infection in the body, including white blood cell (WBC) count, C-reactive protein (CRP) levels and erythrocyte sedimentation rate (ESR).[8] When these tests show evidence of possible infection, patients are referred for further diagnostic testing. The most accurate diagnostic tool is a bone biopsy or aspiration of a pus collection from the bone or tissue surrounding the bone, with a microbiological assessment of the sample to identify the organism causing the infection. Biopsies are invasive and generally require a local or a general anaesthetic. The analysis of the results may take several days. Alternative diagnostic tools include blood or tissue cultures, which may be less accurate but are useful in identifying the organism causing an infection in the body, which enables the selection of the appropriate antibiotic for treatment. The primary treatment for osteomyelitis is a course of antibiotics, but surgery may also be used.[9]

Diagnostic imaging for osteomyelitis

Diagnostic imaging of the affected area before performing a biopsy may help improve diagnosis and avoid unnecessary biopsies in people who may have an infection but are unlikely to have osteomyelitis. It can also be useful to identify pus collections, assess the need for drainage procedures and establish the best way for surgical access.

A range of diagnostic imaging methods are available, including radiography, magnetic resonance imaging (MRI), computed tomography (CT), scintigraphy, positron emission tomography (PET), single-photon emission computed tomography (SPECT) and ultrasound.[9–12] These imaging methods each have their advantages and disadvantages.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Radiography are easily available and cheap to perform, but are poor at detecting osteomyelitis in its early stages. Radiography may be most useful in identifying and ruling out other causes of the patient's symptoms, such as bone fractures.[11] MRI is probably the most widely recommended and used technique. It is more accurate than radiography and able to detect osteomyelitis in its early stages, but is more expensive to perform[12] and when used in children necessitates the use of sedation or general anaesthesia. PET and bone scintigraphy are more expensive and less widely available than MRI or radiography.[11,13] These methods expose patients to ionising radiation. Ultrasound avoids the radiation exposure and is readily available, but its diagnostic accuracy is currently uncertain.[12] There is also a distinction between methods that provide two-dimensional images (radiography, scintigraphy) and those producing three-dimensional images (PET, MRI, CT, SPECT). Some tests (e.g. MRI) may be less suited to patients with hip replacements or other indwelling metalwork because the metalwork can alter the reliability of the imaging.

Current diagnostic and treatment practice

Once osteomyelitis is suspected on the basis of physical examination and blood tests, MRI is currently generally recommended as the imaging test of choice because it can detect osteomyelitis early and it can identify pus collections within bone that might require surgical drainage. Radiography is not usually recommended in isolation, because of their failure to detect early osteomyelitis, but are generally used as a first-line investigation to rule out or confirm bone fractures or other causes of symptoms.[12] CT, scintigraphy and PET are less widely recommended, but are an alternative for patients in whom MRI is not possible.

Ultrasonography is suggested as an alternative to radiological tests[7,9,11,12,14] and is widely used in paediatric practice to exclude joint effusions and pus collection next to bone.[14] This is especially helpful in young children (aged < 6 years), who would require a general anaesthetic for MRI. Ultrasound is also used to guide aspiration and biopsy.

Little formal guidance [such as guidelines produced by the National Institute for Health and Care Excellence (NICE)] exists about which imaging techniques to use to diagnose osteomyelitis. The only current NICE guidance is for the treatment of diabetic foot ulcers.[1] In those patients, radiography is recommended to either confirm advanced osteomyelitis or confirm that the symptoms are due to other causes (e.g. broken bones). Radiography is followed by MRI if osteomyelitis is suspected but not confirmed by radiography. In children, ultrasonography is sometimes used in place of MRI. Antigranulocyte Fab fragment antibody scintigraphy should not be used in patients with diabetic foot ulcers.[15] Recommendations about its use have also been published in the USA.[16,17]

Osteomyelitis is treated with a 4- to 6-week course of antibiotics.[18,19] Treatment is initially intravenous, switching to oral antibiotics after around 2 weeks. The choice of antibiotics will depend on the infecting organism, as determined by tests such as microbiological culture and the patient's medical history. Surgery may also be used for debridement of necrotic tissue and affected bone, to drain pus and to reduce bacterial load.

Pathway to diagnosis in the NHS

There are a number of ways in which a patient might be referred for imaging to diagnose osteomyelitis. Patients may present with fever and be admitted as inpatients, or may be referred directly by their general practitioner (GP) to an orthopaedic clinic. This pathway to clinic is slower than presenting directly to accident and emergency and such patients often have less virulent infection or subacute osteomyelitis. Patients may be referred from other hospitals, particularly those that lack the facilities to treat children (e.g. if the hospital does not offer MRI under general anaesthesia). Patients presenting with acute symptoms may have a musculoskeletal issue (often limping or joint pains) or non-specific systemic symptoms and sepsis (e.g. immunodeficient patients as a result of underlying chronic condition). Generally unwell patients with sepsis are more difficult to diagnose because they might be in intensive care and joint symptoms could initially be missed while the focus is on treating severe symptoms.

The range of symptoms and possible causes of these symptoms mean that osteomyelitis may not be suspected at first. Patients may undergo one or more radiographic assessments (and ultrasound scans in children) and repeated blood tests prior to final diagnosis. Patients may also have received a course of antibiotics before diagnosis, with osteomyelitis suspected only because that treatment course was not successful. This complicates the diagnostic process, and the practical pathway to most diagnoses of osteomyelitis in many patients will differ from that used in formal diagnostic accuracy studies.

Existing review evidence

Several systematic reviews or meta-analyses have been performed to assess diagnostic imaging techniques for osteomyelitis.[20–30]

Four of these reviews considered primarily, or only, people with diabetic foot ulcers.[20,22,23,26] Their conclusions varied, depending on the tests included, but MRI, PET and WBC scintigraphy were all suggested as suitable imaging tests. Three reviews of osteomyelitis in the general population mostly recommended PET and SPECT as having the best diagnostic accuracy.[25,27,28] One review focused on patients with peripheral post-traumatic osteomyelitis and concluded that WBC scintigraphy with SPECT/CT or ^{18}F-FDG (fludeoxyglucose)-PET/CT had the best diagnostic accuracy in this population.[21] Another review of MRI in patients with pressure ulcers was inconclusive as a result of insufficient evidence.[29] Two reviews were conducted in children. One focused on calcaneal osteomyelitis and was inconclusive because of the limited evidence.[24] The other children's review focused on haematogenous acute and subacute paediatric osteomyelitis, and found that MRI had the highest sensitivity and specificity compared with radiography, scintigraphy, CT and ultrasound.[30]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 2 Aims and objectives

The overall aim of this project was to systematically review the literature on diagnostic imaging for osteomyelitis in order to identify the techniques with the best diagnostic accuracy and the greatest clinical utility, across the range of types of disease and patients. The key objectives were to:

- perform a systematic review of all studies reporting the diagnostic accuracy of any relevant imaging test, or combination of tests, used to detect osteomyelitis
- perform diagnostic meta-analyses of identified studies to formally assess their diagnostic accuracy
- investigate diagnostic accuracy across the range of different types of osteomyelitis and types of patient
- compare the diagnostic accuracy of diagnostic tests both statistically and pragmatically, by systematically reviewing inter-rater reliability, and the broader implementation of imaging tests, accounting for key factors such as availability of machinery, radiation exposure and acceptability to patients
- provide useful guidance as to which imaging tests should be preferred, according to type of disease and patient, in the UK.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 3 Methods

A systematic review of the clinical effectiveness was performed following the general principles recommended in the Centre for Reviews and Dissemination's (CRD's) guidance and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. The protocol details have been registered on PROSPERO (number CRD42017068511).

Literature searches

The search strategy was developed by an information specialist with input from the review team and clinical advisors. The strategy was developed in MEDLINE (via Ovid) and included search terms for osteomyelitis and relevant diagnostic imaging techniques. No language, date, geographical or study design limits were applied. The MEDLINE strategy was adapted for use in the other resources searched.

The searches were carried out during August 2017 and updated in July 2018 to capture more recent studies. The following databases were searched: MEDLINE (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE), Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), Cumulative Index to Nursing and Allied Health (CINAHL) Plus, Database of Abstracts of Reviews of Effects (DARE), EMBASE, Health Technology Assessment (HTA) database and PubMed.

In addition, ClinicalTrials.gov and PROSPERO were searched for ongoing and unpublished studies. Relevant guidelines were identified through searches of the National Guidelines Clearing House, NHS Evidence, the NICE website and the Trip database. The reference lists of relevant systematic reviews were manually checked to ensure that all relevant studies from previous reviews were identified.

The search results were imported into EndNote X8 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] and de-duplicated. The complete search strategies can be found in *Appendix 1*.

Study selection

Titles and abstracts and the full texts of studies were independently assessed for inclusion by two reviewers using the inclusion criteria outlined in this section. Disagreements were resolved through discussion and, where necessary, consultation with a third reviewer. Study selection was performed using EPPI-Reviewer 4 software (Evidence for Policy and Practice Information and Co-ordinating Centre, University of London, London, UK).

Participants

Participants were any patients with suspected osteomyelitis (based on symptoms, surgical samples or blood tests). No restrictions were made for age or disease aetiology.

Index tests

Index tests considered included any diagnostic imaging technique that could potentially identify osteomyelitis, either alone or in combination with other relevant tests, such as radiography, MRI, CT, PET, SPECT and ultrasound. Variations on these tests were included, such as variations in the radioisotopes used and differences in protocols or contrast agent use.

Scintigraphy was not a protocol-specified imaging test for this review as it was not expected to be widely used in the UK, particularly because three-dimensional (3D) SPECT imaging may be preferred to planar scintigraphy. However, as the protocol specified that any relevant imaging test would be considered,

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

and many studies of scintigraphy were identified, scintigraphy was included in this review. To focus on the more UK-relevant tests in the main analysis, data on scintigraphy were included only where they were included in a study of another relevant imaging test. Studies evaluating the diagnostic accuracy of scintigraphy alone were considered in a separate analysis.

Reference standards

The preferred reference standard was histopathology or microbiology based on bone biopsy. Surgery was also accepted as a reference standard; other accepted methods of sample collection included pus aspiration.

As biopsies are invasive, clinical follow-up of at least 6 months with no signs or symptoms of osteomyelitis was also accepted as confirmation of the absence of osteomyelitis. Similarly, clinical evidence that the symptoms had another cause was accepted as confirmation of the absence of osteomyelitis.

To avoid potential bias through overestimation of diagnostic accuracy, studies were excluded if a positive osteomyelitis diagnosis was made by using a second imaging test, or by clinical follow-up alone, without biopsy or other microbiological testing.

Outcomes

Diagnostic accuracy review

The primary outcome was the diagnostic accuracy of the imaging test compared with the reference standard expressed in terms of sensitivity (percentage of people/scans with osteomyelitis with a positive diagnostic test result) and specificity (percentage of people/scans without osteomyelitis with a negative test result). Studies reporting sensitivity and specificity, or sufficient data to calculate both measures, were included. Studies evaluating other related conditions, such as spondylodiscitis or septic arthritis, were included if they provided separate diagnostic accuracy data for osteomyelitis.

Inter-rater reliability and implementation of imaging tests

Studies reporting inter-rater reliability for any test, or any other measure of accuracy of test interpretation, were included. Studies reporting information on the broader implementation and acceptability of imaging tests for osteomyelitis were included. The following implementation outcomes were eligible for inclusion: cost-effectiveness of imaging tests (of relevance to the UK), availability of tests (e.g. access to machinery), radiation exposure and substantive data on the experience of patients or clinicians.

Study designs

Any study that considered an imaging test or tests for osteomyelitis that reported data on any of the specified outcomes was included. Therefore, studies reporting any diagnostic accuracy data, other quantitative data (e.g. inter-rater reliability) or substantive qualitative or semiqualitative data (e.g. surveys of patients and clinicians) were included. Only studies explicitly considering testing for osteomyelitis were included. Studies reporting on characteristics of the imaging tests more broadly were excluded.

The following types of reports were excluded: editorials and opinions, case reports, reports focusing only on technical aspects of imaging tests (e.g. technical descriptions or specifications of machinery). We selected the most complete or most recent report in cases of multiple reports for a given study or when we could not exclude the possibility of overlapping populations.

Data extraction

A mapping exercise informed the development of the data extraction form, which was then piloted on a small selection of studies by two reviewers. Data extracted included details of patient characteristics, diagnostic tests and reference standard tests. Data were extracted by one reviewer and independently checked for accuracy by at least one other reviewer. Discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. Given the high number of studies, authors were not contacted if relevant data appeared to be unreported. In cases of multiple reports for a given study, the most recent or most complete report was used as the main source of data.

Study characteristics were extracted, including design, year, country and patient eligibility criteria. Patient characteristics that were extracted included age, comorbidities, diabetic status, location of osteomyelitis and reason for referral. Data on study intervention (e.g. characteristics of imaging test used, radioisotope, contrast agent, diagnostic cut-off point and thresholds), unit of analysis (e.g. patient, body part) and data on exclusions from study/analysis with reasons were recorded. Types of reference standards used for confirming positive and negative cases were recorded. The numbers of patients confirmed to be positive or negative in accordance with the reference standard, and the numbers of true-positive, true-negative, false-positive and false-negative test results, were extracted, if reported. If not reported, sensitivity and specificity estimates [with their 95% confidence intervals (CIs)] or other reported diagnostic accuracy data were extracted. Where possible, sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs) were calculated and checked against reported values in the publications to check for any discrepancies. When more than one test was performed for the same participant in the same body part (i.e. repeat test, or follow-up test), only the result of the first test was used.

Inter-rater reliability estimates were extracted from the papers and tabulated. For the implementation review, relevant results on cost-effectiveness or use of machinery or data from surveys of clinicians were extracted and summarised narratively.

Quality assessment

The quality of the included diagnostic accuracy studies was assessed using the QUADAS-2 [quality assessment of diagnostic accuracy studies (version 2)] tool designed for diagnostic accuracy studies.[31] Critical appraisal was performed by one reviewer alongside data extraction and independently checked by at least one other reviewer. The QUADAS-2 tool was adapted to ensure that it is applicable to assessing the quality of studies of imaging tests for detecting osteomyelitis. The tool consists of four key domains: (1) patient selection, (2) index test, (3) reference standard and (4) flow of patients through the study and timing of the index test(s) and reference standard. Each domain was assessed in terms of the risk of bias. The first three domains were also assessed for concerns regarding their applicability, that is, whether or not the participants and setting, the index test, its conduct or interpretation and the target condition (as defined by the reference standard) were applicable to the review question. Further details on the critical appraisal tool, including signalling questions to inform the assessment of the key domains, are reported in *Appendix 2*.

No validated instrument is available for appraising the quality of studies on inter-rater reliability. We used a modified version of the tool reported by van de Pol *et al.*[32]

Synthesis

Diagnostic meta-analysis

For each diagnostic imaging test, data were synthesised in meta-analyses across studies using logistic regression modelling. This approach fits a statistical model that regresses index test outcome (positive or negative for osteomyelitis) against whether each person does or does not have confirmed osteomyelitis, based on the reference standard. This has been shown[33] to be equivalent to both bivariate meta-analysis and hierarchical summary receiver operating characteristic (HSROC) analysis, which are the methods most commonly used in diagnostic meta-analyses.[34,35] It also accounts for correlation between sensitivity and specificity, and for the fact that these may vary if different test thresholds are used across studies. This proposed model is known as a 'one-stage' approach because it analyses summary diagnostic accuracy across all studies simultaneously. It provides a more flexible approach than conventional bivariate or

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

HSROC analysis; in particular, it permits the inclusion of extra terms in the model to identify subgroups of studies or participants, and to compare different imaging tests.

Studies were pooled if three or more studies were eligible for the analysis. Random-effects models were used to account for potential heterogeneity in diagnostic accuracy across studies. Results were presented as summary sensitivity and specificity estimates, with 95% CIs, plotted in receiver operating characteristic (ROC) space, and as summary HSROC curves.

When studies reported diagnostic accuracy for multiple tests, data for each test were included in the analysis. When studies reported multiple results for the same imaging test (e.g. at different test thresholds, or diagnosis by different clinicians) then only the data corresponding to the greatest diagnostic accuracy [i.e. having the highest diagnostic odds ratio (DOR)] were included.

The PPVs and NPVs were also analysed using the same bivariate logistic regression approach. It should be noted that PPV and NPV depend on the incidence of osteomyelitis in each study, and so may be more heterogeneous than sensitivity and specificity.

In addition to the bivariate analyses, meta-analyses of estimated DOR and positive rates (PRs) (the proportion of people whose imaging test result suggests osteomyelitis, and so who would be diagnosed with osteomyelitis on the basis of the imaging test) were also performed. Univariate meta-analyses (ignoring correlation between outcomes) of sensitivity, specificity, PPV and NPV were also performed, for comparison with the bivariate analyses. In all these meta-analyses, heterogeneity was assessed using I^2.[36]

When the studies were deemed too diverse for meta-analysis to be suitable, or when only one or two studies were available, the reported diagnostic accuracy from each available study was presented in tables and on ROC plots, and compared across studies, tests and subgroups. Studies that did not report full diagnostic accuracy data, but only summary sensitivity and specificity, were summarised in narrative form by tabulating the results.

Subgroup analyses

Separate meta-analyses were conducted for each diagnostic imaging test and, where sufficient data were available, according to the following subcategories of patients:

- children and adults (as defined by the studies)
- cause of osteomyelitis (haematogenous, contiguous, trauma, surgical, diabetes related, other)
- acute, subacute or chronic osteomyelitis (COM)
- anatomical site (long bone, spinal, foot and ankle, pelvis, other)
- patients with hip replacements or other indwelling metalwork.

Subgroup analyses were also performed to assess the impact of different study characteristics:

- subtypes of imaging test (e.g. owing to the use of different radioisotopes, inclusion or exclusion of CT scanning, different thresholds, or methods of image interpretation)
- choice of reference standard (biopsy, or clinical and surgical follow-up)
- study quality (based on QUADAS-2 criteria).

As having diabetic foot ulcers is one of the most common causes of osteomyelitis, we performed meta-analyses specifically of studies of people with diabetic foot ulcers.

Analyses were performed separately for adults and children.

Analyses within subgroups were performed using the logistic regression analysis approach discussed in *Diagnostic meta-analysis*. Separate analyses were performed for each subgroup.

Comparison of imaging tests

Diagnostic tests were compared by examining summary DORs derived from the logistic regression models and by comparing summary ROC curves. In general, a larger DOR indicates a better performance, but this may not be the case if ROC curves cross, in which case the trade-off between sensitivity and specificity was considered. When there were sufficient data, these comparisons were made in each of the subgroups listed in *Subgroup analyses*.

When studies reported diagnostic accuracy data for two or more imaging tests on the same patient population, these tests were compared within the study by comparing sensitivity, specificity and DOR estimates. The main bivariate logistic regression models were extended to include all imaging tests in one model, to allow tests to be formally compared for differences in DOR and specificity.

Inter-rater reliability and implementation review

Owing to heterogeneity, studies evaluating inter-rater reliability results were reported narratively and tabulated. For studies reporting qualitative data on the implementation of diagnostic tests (e.g. clinical or patient opinions), data were synthesised using a narrative synthesis approach. Areas where few or no data have been published were also identified.

Deviations from the protocol

Some changes in the review and analysis process were made as a consequence of the nature of the identified studies.

Scintigraphy was not a protocol-specified imaging test, but many of the included studies compared scintigraphy with another test. Scintigraphy was therefore included in the main review and meta-analyses where studies compared it with another eligible test. To ensure a fair analysis of scintigraphy, a separate meta-analysis of scintigraphy alone was performed. In that analysis, we also included studies reporting the diagnostic accuracy of scintigraphy alone that included > 20 participants and that had been published since 1990. These additional studies were not assessed for quality using QUADAS-2 and only basic diagnostic accuracy data were extracted from publications.

For several proposed subgroup analyses, either we identified no relevant studies or studies did not report sufficient data for synthesis, so no analyses were performed. The subgroups were patients with:

- acute symptoms (such as would be admitted as inpatients)
- sepsis
- milder or chronic symptoms (such as would be referred by a GP)
- concomitant diseases (e.g. cancer).

The number of studies of children and of ultrasound were too few to permit a meta-analysis, so these studies were combined in a narrative synthesis only.

Role of patient and clinical advisors

Clinical advisors for this project attended meetings (in person or remotely) over the course of the project to ensure that it met clinical needs. This included approving the protocol, discussing and commenting on the provisional results and meta-analyses, and commenting on this report.

Two patient representatives, both parents of children who had been treated for osteomyelitis, were contacted through telephone meetings over the course of the project to discuss their experience of imaging tests and to discuss the findings of the project.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Chapter 4 Results

Quantity and quality of research available

Included studies

Figure 1 is a flow diagram outlining the screening process with reasons for exclusion of full-text papers. The literature searches of bibliographic databases identified 18,386 unique references. After initial screening of titles and abstracts, 597 were considered to be potentially relevant and were ordered for full-paper screening. In total, 77 studies[37–113] (from 79 reports) were included in the diagnostic review, of which 69 were also included in a meta-analysis.[38–42,44–51,53–71,73–77,79–95,97,99–113] Eleven studies were included in the review of inter-rater agreement,[37,58,59,65,68,86,92,114–117] and one was included in the review of implementation.[118] Seven studies were included in both the diagnostic and inter-rater reliability reviews.[37,58,59,65,68,86,92]

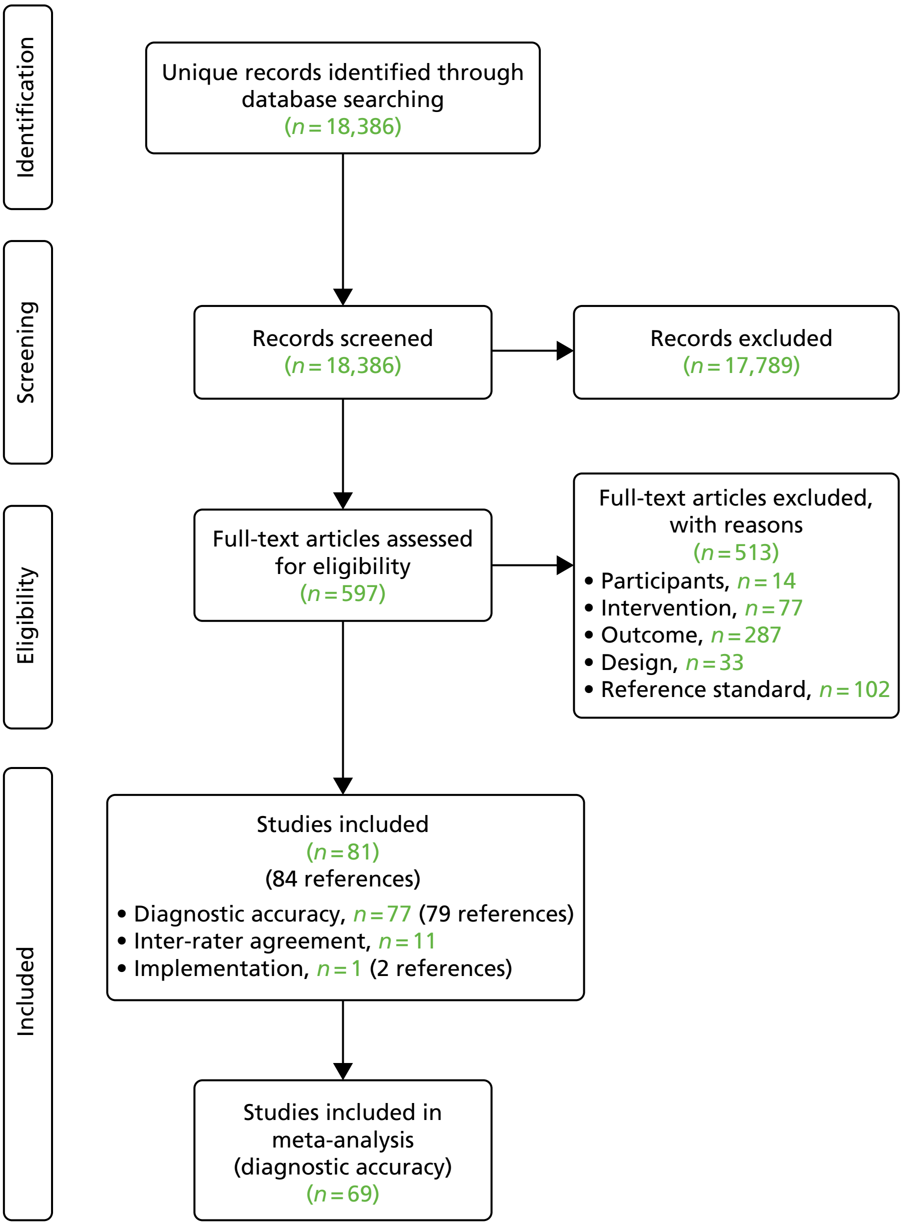

Identification
Unique records identified through database searching (n = 18,386)
Screening
Records screened (n = 18,386)
Records excluded (n = 17,789)
Eligibility
Full-text articles assessed for eligibility (n = 597)
Full-text articles excluded, with reasons (n = 513)
• Participants, n = 14
• Intervention, n = 77
• Outcome, n = 287
• Design, n = 33
• Reference standard, n = 102
Included
Studies included (n = 81) (84 references)
• Diagnostic accuracy, n = 77 (79 references)
• Inter-rater agreement, n = 11
• Implementation, n = 1 (2 references)
Studies included in meta-analysis (diagnostic accuracy) (n = 69)

FIGURE 1 A PRISMA flow diagram showing study selection.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Excluded studies

Of the 597 potentially relevant studies, the most common reason for exclusion was lack of relevant outcome data (287 references, of which 49 did not have a full text available) and 102 references did not report an eligible reference standard. Fifty-four studies[69,119–171] evaluated the diagnostic accuracy of scintigraphy but were excluded from the main review because they did not evaluate the diagnostic accuracy of other eligible index tests. The list of references that were excluded along with the reasons for their exclusions is not reported because of the large number of exclusions, but is available from the authors on request.

Assessment of diagnostic accuracy

Characteristics of included studies

Table 1 presents a summary of design and patient characteristics of all included studies. Most studies (46 out of 77) were published in or after 2000.[37,38,40,41,43–46,48,51–57,60–65,67–70,72,73,77,78,80–82,84,88,93,95–99,103,105,108,109,111,113]

Just under half of the studies were conducted in the USA (34 studies).[39,42,46,56,63,66,67,70–72,74–76,79,80,82,83,85,86,88–92,98,100–102,104,106,107,110–112] Eight studies were conducted in Germany,[49,58–60,64,65,81,108] six in Switzerland,[41,44,61,68,73,99] five in Italy[50,53–55,95] and three in Egypt,[37,52,113] the Netherlands[48,62,105] and Turkey.[51,96,97] Two studies were conducted in France,[41,45] India,[77,93] the Republic of Korea,[69,109] Spain[40,84] and the UK.[41,94] A single study was conducted in the following countries: Austria,[108] Belgium,[57] Bulgaria,[43] Canada,[43] Israel,[103] Japan,[82] the Republic of Korea,[109] Kuwait,[38] Lithuania[78] and Oman.[87] Three studies were conducted in more than one country.[41,82,108]

None of the included studies was randomised, and 41 studies used a prospective design.[37,41–43,45,47,49,51,52,54,55,57–60,62,64–66,75–78,81,83–85,88–91,93,94,96,99,104,107,110–113]

The sample size of the studies ranged from 7 to 339, but most studies were small: only 15 (20%) included ≥ 50 participants.[40,41,46,50,67,70,71,75,78,84,86,88,98,108,113] Nearly all studies were conducted in adults. Three studies were conducted in children exclusively,[52,78,98] and 10 were conducted in a mixed population of adults and children.[44,50,57,85,87,96,101,104,106,109] Nine studies included at least some participants with indwelling work, [39,57,61,69,73,79,81,103,108,109] but only one focused exclusively on this population.[103]

Nearly half of the included studies included only or mostly patients with diabetes (36 studies).[37,38,40–43,47,49,51,53,55,56,62,67,70,71,74,76,77,82,84,86,88–90,92–95,97,99,102,107,111–113] A total of 10 studies focused on patients with trauma and/or previous surgery.[57,59,61,64,68,69,72,73,79,103,105,108] Osteomyelitis was caused by non-foot-related pressure ulcers and skin ulcerations in six studies[45,63,66,72,75,80] and one study attributed osteomyelitis to each of the following conditions specifically: soft tissue infection,[96] septic arthritis[98] and neuropathic osteoarthropathy.[100] Eighteen studies included patients with multiple aetiologies,[44,46,48,50,52,54,58,60,65,78,81,83,85,91,101,106,109,110] and three studies did not report any specific causes.[39,87,104] Only one study included patients with haematogenous osteomyelitis.[78] *Appendix 3* presents further details on the study participant inclusion and exclusion criteria.

Forty-seven studies stated which type of osteomyelitis they targeted. Twelve studies aimed to detect acute osteomyelitis (AOM) specifically,[37,51,52,66,74,78,87,90,91,98,99,104] 17 focused on COM[46,49,58,60,61,64,68,73,79,81,93,94,96,108,111–113] and 15 focused on a mixed population of AOM and COM.[38,42,44,48,50,54,57,63,65,69,71,80,85,92,106,107] Two studies targeted patients with either subacute osteomyelitis or COM.[39,59] The condition targeted was not specified in 30 studies.[40,41,43,45,47,53,55,56,62,67,70,72,75–77,82–84,86,88,89,95,97,100–103,105,109,110] Most studies used the dual reference standard for osteomyelitis of clinical follow-up for patients who tested negative and bone biopsy with histopathology or microbiology for test-positive patients, although 34 studies used histopathology/microbiology in all participants regardless of their imaging test results.[37,38,42,43,45,49,51,53,54,56–59,61,62,70,72,75,77,79–82,84,89,90,92,93,98,100,102,107,108,113]

Table 2 presents the distribution of included diagnostic accuracy studies according to tests and body part targeted. This shows that most of the evidence focused on the diagnostic accuracy of MRI, scintigraphy and radiography for the diagnosis of diabetic foot osteomyelitis. Few studies specifically focused on the axial skeleton, the pelvis/hip/knee and long bones.

TABLE 1 Study design and participant characteristics

First author and year of publication	Country	Design	*n*	Population	Mean (SD) age [range] (years)	% male	Cause of osteomyelitis	Indwelling work (%)	Body part targeted
Abdel Razek (2017)[37]	Egypt	Prospective	39	Adults	51 [48–72]	54	Diabetic foot ulcer	NR	Foot
Al-Khawari (2005)[38]	Kuwait	Retrospective	19	Adults	61 [41–81]	59	Diabetic foot ulcer	NR	Foot/ankle
Al-Sheikh (1985)[39]	USA	NR	22	Adults	NR [32–65]	NR	Multiple (NR)	48	Multiple
Aragón-Sánchez (2011)[40]	Spain	Retrospective	338	Adults	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Aslangul (2013)[41]	France, UK and Switzerland	Prospective	53	Adults	63 (10) [NR]	93	Diabetic foot ulcer	NR	Foot
Blume (1997)[42]	USA	Prospective	27	Adults	64.7 (1) [NR]	96	Diabetic foot ulcer (93%), other (7%)	NR	Foot
Bohchelian (2002)[43]	Bulgaria	Prospective	32	Adults	NR	NR	Diabetic foot ulcer	NR	Foot
Bolouri (2013)[44]	Switzerland	Retrospective	42	Mixed	52 [10–84]	50	Multiple	NR	Jaw/head
Brunel (2016)[45]	France	Prospective	34	Adults	51 [NR]	71	Pressure ulcer	0	Pelvis/hip
Chacko (2003)[46]	USA	Retrospective	56	NR	NR [NR]	NR	Multiple	0	Multiple
Croll (1996)[47]	Canada	Prospective	27	Adults	66 [34–82]	70	Diabetic foot ulcers	NR	Foot
Demirev (2014)[48]	The Netherlands	Retrospective	26	Adults	59 [16–78]	54	Multiple	NR	Multiple
Enderle (1999)[49]	Germany	Prospective	19	Adults	61 (10) [NR]	90	Diabetic foot ulcers	NR	Foot
Erdman (1991)[50]	Italy	Retrospective	110	Mixed	37 [0–74]	71	Multiple	0	Multiple
Ertugrul (2006)[51]	Turkey	Prospective	26	NR	62 (9) [40–77]	74	Diabetic foot ulcer	NR	Foot
Ezzat (2011)[52]	Egypt	Prospective	27	Children	10 [NR]	67	Multiple	NR	NR
Familiari (2011)[53]	Italy	NR	13	Adults	62 [50–89]	92	Diabetic foot ulcers	NR	Foot
Filippi (2006)[54]	Italy	Prospective	15	Adults	50 (13) [NR]	61	Multiple	0	Multiple
Filippi (2009)[55]	Italy	Prospective	17	Adults	55 (4) [NR]	59	Diabetic foot ulcer	NR	Foot
Franceschi (2013)[56]	USA	NR	17	Adults	NR [21–84]	62	Diabetic foot ulcer	NR	Foot

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 1 Study design and participant characteristics (*continued*)

First author and year of publication	Country	Design	*n*	Population	Mean (SD) age [range] (years)	% male	Cause of osteomyelitis	Indwelling work (%)	Body part targeted
Gemmel (2004)[57]	Belgium	Prospective	22	Mixed	40 [14–70]	45	Surgical	82	Spine
Guhlmann (1998)[58]	Germany	Prospective	48	Adults	49 [22–81]	76	Trauma or surgery	NR	Multiple
Guhlmann (1998)[59]	Germany	Prospective	31	Adults	48 [20–78]	74	Trauma or surgery	NR	Multiple
Hakim (2006)[60]	Germany	Prospective	34	Adults	54 (19) [19–93]	NR	Dental infection, trauma, others	NR	Jaw
Hartmann (2007)[61]	Switzerland	Retrospective	33	Adults	50 [17–80]	79	Trauma	55	Multiple
Hazenberg (2011)[62]	The Netherlands	Prospective	21	NR	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Heiba (2017)[63]	USA	Retrospective	33	Adults	NR	58	Pelvic pressure ulcer	NR	Pelvis/hip
Horger (2003)[64]	Germany	Prospective	27	Adults	48 [20–90]	82	Trauma	NR	Multiple
Horger (2007)[65]	Germany	Prospective	31	Adults	51 [20–89]	52	Multiple (musculoskeletal infection)	NR	Multiple
Huang (1998)[66]	USA	Prospective	42	Adults	41 [22–83]	82	Paralysis and skin ulceration	NR	Pelvis/hip
Johnson (2009)[67]	USA	Retrospective	73	Adults	65 [26–92]	66	Diabetes, neuropathic arthropathy, other	NR	Foot
Kaim (2000)[68]	Switzerland	Retrospective	18	Adults	45 [27–65]	72	Trauma	0	Long bone
Kim (2017)[69]	Republic of Korea	Retrospective	21	Adults	51.9 (18.1) [23–77]	67	Trauma	48	Lower extremities
La Fontaine (2016)[70]	USA	Retrospective	52	Adults	50 (10) [26–74]	73	Diabetes and peripheral neuropathy	NR	Foot
Larcos (1991)[71]	USA	Retrospective	51	Adults	62 [30–88]	61	Diabetic foot ulcers	NR	Foot
Larson (2011)[72]	USA	Retrospective	44	Adults	NR [NR]	NR	Pelvic pressure ulcer	NR	Pelvis/hip
Ledermann (2000)[73]	Switzerland	Retrospective	15	Adults	41 [26–66]	67	Trauma	73	Lower extremities
Levine (1994)[74]	USA	Retrospective	27	Adults	52 [33–72]	44	Diabetic foot ulcer	NR	Foot
Lewis (1988)[75]	USA	Prospective	52	Adults	NR [NR]	NR	Spinal cord injury and skin ulceration	NR	Pelvis/hip

First author and year of publication	Country	Design	*n*	Population	Mean (SD) age [range] (years)	% male	Cause of osteomyelitis	Indwelling work (%)	Body part targeted
Lipman (1998)[76]	USA	Prospective	20	Adults	46 [28–72]	65	Diabetes (85%), peripheral neuropathy (100%)	NR	Foot
Mahendra (2017)[77]	India	Prospective	34	Adults	52 (9) [NR]	65	Diabetic foot ulcer	0	Foot
Malcius (2009)[78]	Lithuania	Prospective	169	Children	10 (4) [1–18]	69	Multiple (haematogenous)	NR	NR
Mason (1989)[79]	USA	Retrospective	14	NR	NR [NR]	NR	Trauma (93%), other (7%)	14	Long bone
McCarthy (2017)[80]	USA	Retrospective	41	Adults	45 [NR]	82	Skin ulceration	NR	Multiple
Meller (2002)[81]	Germany	Prospective	30	Adults	NR [24–72]	47	Multiple	13	Multiple
Miki (2015)[82]	Japan and the USA	Retrospective	26	Adults	67 [42–85]	77	Diabetic foot ulcers	NR	Foot
Modic (1985)[83]	USA	Prospective	37	Adults	53 [22–85]	51	Multiple	NR	Spine
Morales Lozano (2010)[84]	Spain	Prospective	132	NR	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Morrison (1993)[85]	USA	Prospective	49	Mixed	49 [10–80]	NR	Multiple	NR	Multiple
Morrison (1998)[86]	USA	Retrospective	68	Adults	56 [24–85]	NR	Diabetes (85%), others (15%)	NR	Foot
Nath (1992)[87]	Oman	NR	25	Mixed	NR [2–45]	NR	NR	NR	Long bone
Nawaz (2010)[88]	USA	Prospective	106	Adults	59 [29–85]	69	Diabetic foot ulcer	NR	Foot
Newman (1991)[89]	USA	Prospective	35	NR	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Newman (1992)[90]	USA	Prospective	12	NR	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Nigro (1992)[91]	USA	Prospective	44	Adults	55 [23–84]	57	Multiple (diabetic foot ulcer, foot ulcer or inflammation)	0	Foot
Park (1982)[92]	USA	Retrospective	36	NR	NR [NR]	NR	Diabetic foot ulcer	NR	Foot
Rastogi (2016)[93]	India	Prospective	23	Adults	58 (8) [NR]	96	Diabetes, retinopathy, neuropathy and/or vascular disease	0	Foot
Remedios (1998)[94]	UK	Prospective	9	Adults	57 [25–70]	45	Diabetic foot ulcer	NR	Foot
Rozzanigo (2009)[95]	Italy	Retrospective	16	Adults	58 [42–78]	69	Diabetic foot ulcers	NR	Foot
Şanlı (2011)[96]	Turkey	Prospective	30	Mixed	52 [10–93]	71	Soft tissue infection	0	Multiple

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 1 Study design and participant characteristics (*continued*)

First author and year of publication	Country	Design	*n*	Population	Mean (SD) age [range] (years)	% male	Cause of osteomyelitis	Indwelling work (%)	Body part targeted
Sarikaya (2003)[97]	Turkey	NR	26	Adults	59 [18–80]	77	Diabetic foot ulcer	NR	Foot
Schlung (2016)[98]	USA	Retrospective	54	Children	6 (4) [NR]	59	Septic arthritis	NR	Pelvis/hip
Schwegler (2008)[99]	Switzerland	Prospective	20	Adults	66 [53–89]	60	Diabetic foot ulcer	NR	Foot
Seabold (1990)[100]	USA	Retrospective	14	Adults	NR [24–68]	64	Neuropathic osteoarthropathy. Diabetes (78%), others (22%)	NR	Foot (71%), others
Seabold (1995)[101]	USA	Retrospective	26	Mixed	55 [3–78]	52	Multiple	NR	Jaw/head
Segall (1989)[102]	USA	Retrospective	23	Adults	58 (13) [25–79]	95	Diabetic foot ulcer	NR	Foot
Shemesh (2015)[103]	Israel	Retrospective	10	Adults	35 [18–53]	80	Trauma or surgery	100	Long bone
Unger (1988)[104]	USA	Prospective	35	Mixed	52 [1–84]	57	Diabetes (47%), others (NR) (53%)	NR	Multiple
van Vliet (2018)[105]	The Netherlands	Retrospective	30	Adults	46 [18–74]	70.0	Trauma and septic delayed union	NR	Long bone (leg), heel
Weber (1995)[106]	USA	Retrospective	20	Mixed	46 [3–74]	50	Multiple	NR	Jaw/head
Weinstein (1993)[107]	USA	Prospective	32	Adults	49 [23–81]	68	Diabetic foot infections	NR	Foot
Wenter (2016)[108]	Germany and Austria	Retrospective	131	Adults	47 [NR]	82	Trauma	28	Multiple
Weon (2000)[109]	Republic of Korea	Retrospective	37	Mixed	44 [6–77]	68	Multiple	68	Hip and/or knee
Williamson (1989)[110]	USA	Prospective	7	Adults	57 [40–71]	43	Diabetes (57%), others (43%)	0	Foot
Yang (2016)[111]	USA	Prospective	48	Adults	60 (15) [36–83]	67	Diabetic foot ulcer	NR	Foot
Yuh (1989)[112]	USA	Prospective	24	Adults	58 [32–74]	NR	Diabetic foot ulcer	NR	Foot
Zaiton (2014)[113]	Egypt	Prospective	102	Adults	52 (6) [NR]	41	Diabetic foot ulcer	NR	Foot

NR, not reported; SD, standard deviation.

TABLE 2 Diagnostic accuracy studies according to test and body part targeted

	Test (number of studies)						
Body part targeted	CT	MRI	PET and PET/CT	Scintigraphy	SPECT and SPECT/CT	Ultrasound	Radiography
Axial (jaw/head/spine)	3[44,101,106]	2[83,101]	1[60]	3[44,57,83]	5[44,57,60,101,106]	0	2[44,83]
Foot	1[110]	24[37,38,47,49,51,62,67,70,74,76,77,82,86,88,90,91,93–95,99,100,107,112,113]	6[53,56,88,93,99,111]	19[42,47,49,51,53,71,89–92,94,97,99,100,102,107,110,112,136]	3[41,55,70]	1[49]	17[40,42,47,49,71,74,76,84,88,89,91,92,95,97,102,107,112]
Pelvis/hip/knee	2[72,75]	3[45,66,98]	0	3[63,75,109]	2[63,109]	0	2[72,75]
Long bone	0	2[68,79]	1[103]	2[68,79]	0	1[87]	1[87]
Multiple/NR	1[78]	7[48,50,73,78,80,85,104]	8[46,48,58,59,61,81,105,108]	7[39,58,74,78,81,85,104]	4[54,64,65,69]	2[52,78]	2[39,78]

NR, not reported.

Nearly all studies interpreted images by visual assessment without any quantitative analysis (e.g. measuring the standardised uptake value). Some studies compared the results of qualitative and quantitative analysis in parallel, including PET (two studies)[53,108] and WBC scintigraphy.[53] Other studies used a semiquantitative approach (combined qualitative and quantitative interpretation) for PET/CT (two studies),[105,108] PET[108] scintigraphy + SPECT[64] and SPECT/CT.[64]

Critical appraisal

Table 3 presents a summary of the results of the critical appraisal. Further details including justifications for decisions are presented in *Appendix 2*. Overall, 23% of studies were rated as being at a high risk of bias for at least one domain,[39,47,48,52,56,57,64,65,78,80,83,91,92,95,101,102,107,111] although poor reporting of study methods, particularly regarding the selection of patients and the conduct of the index test, means that there is significant uncertainty about the quality of most of the studies.

TABLE 3 Summary of critical appraisal of diagnostic accuracy studies

	Risk of bias				Applicability		
First author and year of publication	Patient selection	Index test	Reference standard	Patient flow	Patient selection	Index test	Reference standard
Abdel Razek (2017)[37]	?	+	+	+	+	+	+
Al-Khawari (2005)[38]	?	+	+	+	+	+	+
Al-Sheikh (1985)[39]	+	–	+	+	+	+	+
Aragón-Sánchez (2011)[40]	+	+	+	+	+	+	+
Aslangul (2013)[41]	+	+	+	+	+	+	+
Blume (1997)[42]	?	+	+	+	+	+	+
Bohchelian (2002)[43]	?	+	+	+	+	+	+
Bolouri (2013)[44]	?	?	+	+	+	+	+
Brunel (2016)[45]	?	+	+	+	+	+	+
Chacko (2003)[46]	?	+	+	+	+	+	+
Croll (1996)[47]	?	–	+	+	+	+	+
Demirev (2014)[48]	+	–	+	+	+	+	+

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 3 Summary of critical appraisal of diagnostic accuracy studies (*continued*)

First author and year of publication	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Patient flow	Patient selection	Index test	Reference standard
Enderle (1999)[49]	+	+	+	+	+	+	+
Erdman (1991)[50]	+	+	+	+	+	+	+
Ertugrul (2006)[51]	?	?	+	+	+	+	+
Ezzat (2011)[52]	–	+	?	+	+	+	+
Familiari (2011)[53]	?	+	+	+	+	+	+
Filippi (2006)[54]	+	+	+	+	+	+	+
Filippi (2009)[55]	+	+	+	+	+	+	+
Franceschi (2013)[56]	?	+	+	–	?	+	+
Gemmel (2004)[57]	–	+	+	+	?	+	+
Guhlmann (1998)[58]	?	+	+	+	+	+	+
Guhlmann (1998)[59]	?	+	+	+	+	+	+
Hakim (2006)[60]	?	?	+	+	+	+	+
Hartmann (2007)[61]	?	+	+	+	+	+	+
Hazenberg (2011)[62]	?	?	+	+	+	+	+
Heiba (2017)[63]	?	+	+	+	+	+	+
Horger (2003)[64]	?	–	+	+	+	+	+
Horger (2007)[65]	+	–	+	+	+	+	+
Huang (1998)[66]	+	+	+	+	+	+	+
Johnson (2009)[67]	?	+	+	+	+	+	+
Kaim (2000)[68]	?	+	+	+	+	+	+
Kim (2017)[69]	+	+	+	+	+	+	+
La Fontaine (2016)[70]	?	+	+	+	+	+	+
Larcos (1991)[71]	?	?	+	+	+	+	+
Larson (2011)[72]	?	?	+	+	?	+	+
Ledermann (2000)[73]	+	?	+	+	+	?	+
Levine (1994)[74]	?	?	+	+	+	+	+
Lewis (1988)[75]	?	?	+	+	?	+	+
Lipman (1998)[76]	+	+	+	?	+	+	+
Mahendra (2017)[77]	+	+	+	+	+	?	+
Malcius (2009)[78]	?	–	?	?	+	+	+
Mason (1989)[79]	+	?	+	+	+	+	+
McCarthy (2017)[80]	–	+	+	+	+	+	+
Meller (2002)[81]	+	+	+	+	+	+	+
Miki (2015)[82]	?	?	+	+	+	+	+
Modic (1985)[83]	?	–	+	+	?	+	+
Morales Lozano (2010)[84]	?	?	+	+	+	+	+
Morrison (1993)[85]	?	?	+	+	+	+	+
Morrison (1998)[86]	?	+	+	+	+	+	+
Nath (1992)[87]	?	?	+	+	+	+	+

TABLE 3 Summary of critical appraisal of diagnostic accuracy studies (*continued*)

	Risk of bias				Applicability		
First author and year of publication	Patient selection	Index test	Reference standard	Patient flow	Patient selection	Index test	Reference standard
Nawaz (2010)[88]	+	?	+	+	+	+	+
Newman (1991)[89]	+	?	+	+	+	+	+
Newman (1992)[90]	?	+	+	+	+	+	+
Nigro (1992)[91]	+	–	?	+	+	+	+
Park (1982)[92]	–	+	+	+	+	+	+
Rastogi (2016)[93]	+	?	+	+	+	+	+
Remedios (1998)[94]	?	+	+	+	+	+	+
Rozzanigo (2009)[95]	?	–	?	?	+	+	+
Şanlı (2011)[96]	+	+	?	?	+	+	+
Sarikaya (2003)[97]	?	+	+	+	+	+	+
Schlung (2016)[98]	?	+	+	+	+	+	+
Schwegler (2008)[99]	+	+	+	+	+	+	+
Seabold (1990)[100]	?	+	+	?	+	+	+
Seabold (1995)[101]	?	+	+	–	?	+	+
Segall (1989)[102]	–	+	+	+	+	+	+
Shemesh (2015)[103]	?	+	+	+	+	+	+
Unger (1988)[104]	+	+	?	?	+	+	+
van Vliet (2018)[105]	+	?	+	+	+	+	+
Weber (1995)[106]	?	+	+	+	+	+	+
Weinstein (1993)[107]	+	+	+	–	+	+	+
Wenter (2016)[108]	?	+	+	+	+	+	+
Weon (2000)[109]	+	+	+	+	+	+	+
Williamson (1989)[110]	?	?	+	+	+	+	+
Yang (2016)[111]	+	+	–	–	+	+	+
Yuh (1989)[112]	+	+	+	+	+	+	+
Zaiton (2014)[113]	+	+	+	+	+	+	+

+, low risk of bias; –, high risk of bias; ?, unclear risk of bias.

Over half of the studies (56%) did not provide sufficient information to assess the risk of bias associated with the selection and enrolment of patients into the study.[37,38,42–47,51,53,56,58–64,67,68,70–72,74,75,78,82–87,89,94,95,97,98,100,101,103,106,108,110] In particular, most studies did not explicitly state whether or not all eligible patients were included during a defined period of time. Therefore, the risk of selection bias, for instance because of the exclusion of patients who are harder to diagnose, cannot be excluded. Nine studies did not blind the interpretation of the index test to the results of other index tests or to the reference standard, and were therefore considered as being at a high risk of bias.[39,47,48,64,65,78,83,91,95] Another 18 studies did not provide sufficient information on blinding and were therefore considered as being at an unclear risk of bias.[44,51,60,62,71–75,79,82,84,85,87,88,90,93,110] Nearly all studies were rated as being at a low risk of bias associated with the reference standard (91%) and patient flow (87%) and none of the included studies raised significant concerns about their applicability to the diagnostic accuracy review questions.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Synthesis of diagnostic accuracy in adults

In this section we consider the synthesis of diagnostic accuracy for studies of imaging tests in adults. Studies exclusively in children are considered in *Synthesis of studies in children*. Studies in mixed adult/children groups or where the population was unclear are included in the analysis of adults.

Excluded from this analysis are studies where the combined diagnostic accuracy of two or more imaging tests was reported (e.g. the accuracy of scintigraphy in combination with SPECT); these are discussed in *Studies reporting on combinations of imaging tests*. Also excluded are studies that did not report sufficient data to calculate 2 × 2 tables of diagnostic accuracy; these are discussed in *Studies not included in the quantitative synthesis*.

This analysis included 69 diagnostic accuracy studies.[38–42,44–51,53–71,73–77,79–95,97,99–113] *Appendix 4*, *Table 19*, summarises these studies. Many of these studies compare two or more imaging tests. Comparisons between tests are discussed in *Comparisons between tests*. Many studies compare imaging tests with scintigraphy. Scintigraphy was not specified as a test of interest in the protocol, but it is included here because of the number of studies considering it. The main analysis only considers scintigraphy in studies where it was compared with other tests; for studies of scintigraphy alone, see *Studies of scintigraphy*.

Some studies reported diagnostic accuracy at multiple thresholds of a single test, or for a single test under different conditions. Where that was the case, the main analysis uses only the result with the highest DOR, as an indicator of the 'best' diagnostic accuracy for that test in that study. Where appropriate, data at multiple thresholds are considered in the sections on individual tests (see *Synthesis of specific imaging tests*).

Figure 2 summarises the sensitivity (proportion of people/scans with osteomyelitis correctly diagnosed) and specificity (proportion of people/scans without osteomyelitis correctly diagnosed) for each test in each study in a ROC plot. This figure shows considerable variation in the diagnostic accuracy both within and between tests. MRI and SPECT both generally have high sensitivity (> 80%), but with a wide range of specificities. PET is generally more consistent with high sensitivity and specificity, but with some studies having lower sensitivity. Scintigraphy results vary widely in both sensitivity and specificity. Radiographic results generally have poor sensitivity.

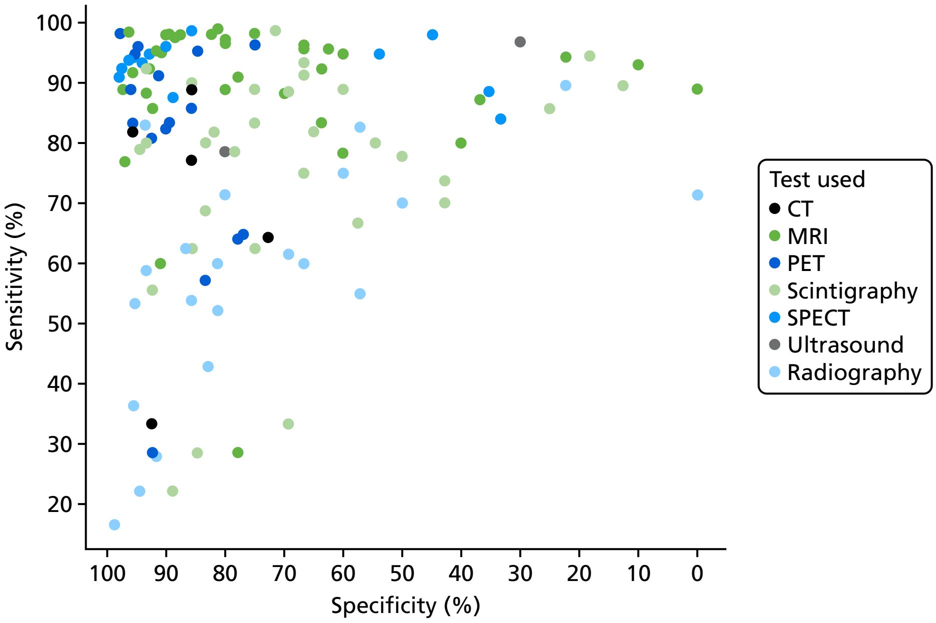


FIGURE 2 Summary ROC plot of diagnostic accuracy in adults.

In *Appendix 4*, *Figure 21* shows a similar plot for PPVs and NPVs. These results are also variable and difficult to interpret. MRI, SPECT and PET generally have high PPV and NPV, but with many outliers. Radiography generally has a poorer NPV.

We now consider the bivariate meta-analysis of sensitivity and specificity. *Figure 3* shows the summary bivariate meta-analysis results for each test (with 95% CIs for sensitivity and specificity). *Figure 4* shows the summary HSROC curves for each test.

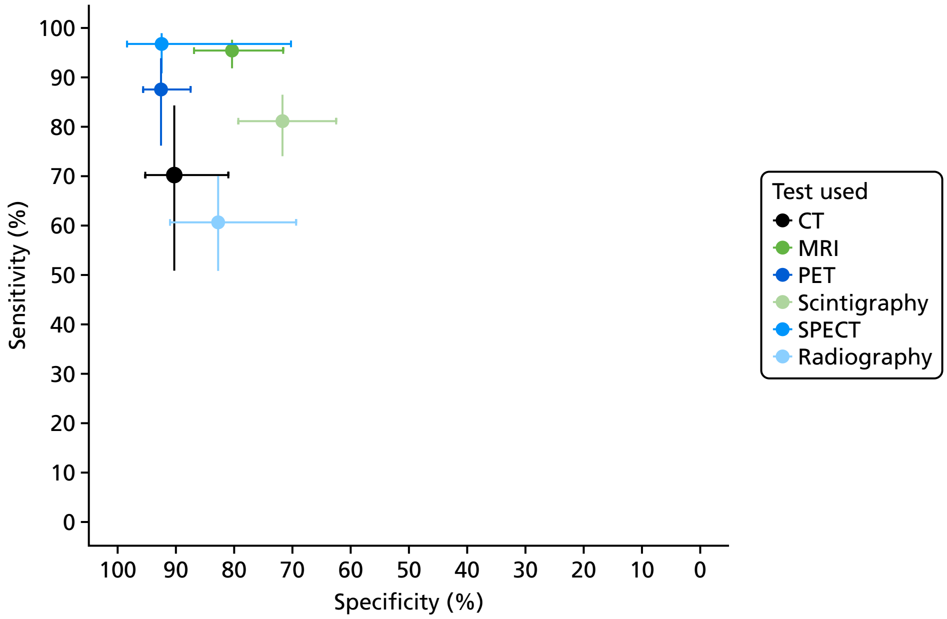


FIGURE 3 Bivariate meta-analysis of sensitivity and specificity: all adult studies.

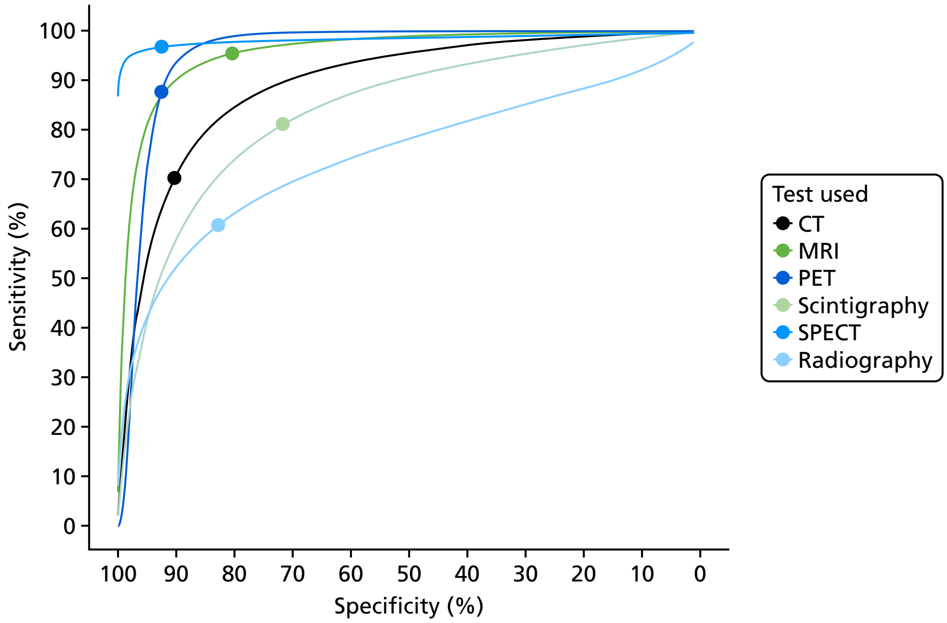


FIGURE 4 Summary HSROC curves: all adult studies.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

These results show that sensitivity and specificity are highest for SPECT, but with a wide CI for specificity. MRI and PET also have high sensitivity and specificity. CT, radiography and scintigraphy all have poorer sensitivity and/or specificity. The summary HSROC curves for MRI, SPECT and PET are very similar and overlapping, suggesting that they have similar diagnostic accuracy. The curves for CT, radiography and scintigraphy are lower, showing poorer diagnostic accuracy.

Appendix 4, *Figure 22*, gives the results of the bivariate meta-analysis of PPV and NPV, which shows that MRI, PET and SPECT have higher PPV (88–93%) and NPV (> 85%) than MRI and PET, with SPECT having the highest NPV (97%). Scintigraphy and radiography have lower PPV and NPV.

Figure 5 shows forest plots of sensitivity and specificity for MRI. These highlight the generally high sensitivity and specificity of MRI, and the overall consistency of sensitivity across studies. As noted earlier, the specificity of MRI varies substantially across studies, although many studies are small with correspondingly wide CIs. This raises concerns as to the consistency of MRI performance across studies. Given the large number of tests, and hence forest plots, no further forest plots are presented in this report, but they are available on request from the authors.

Table 4 shows the results of univariate meta-analyses of sensitivity and specificity (i.e. meta-analysed separately rather than jointly) and also meta-analyses of DOR (to summarise overall diagnostic accuracy) and the PR (proportion of people testing positive for osteomyelitis based on the imaging test result alone).

The results for sensitivity and specificity match those of the bivariate analysis in *Figure 3*, as would be expected. The I^2 values for sensitivity and specificity are low or zero, which conflicts with the apparent heterogeneity observed in *Figure 2*. This may be because most studies are small, and so there is considerable within-study variability in estimates, which reduces I^2, rather than an absence of heterogeneity.

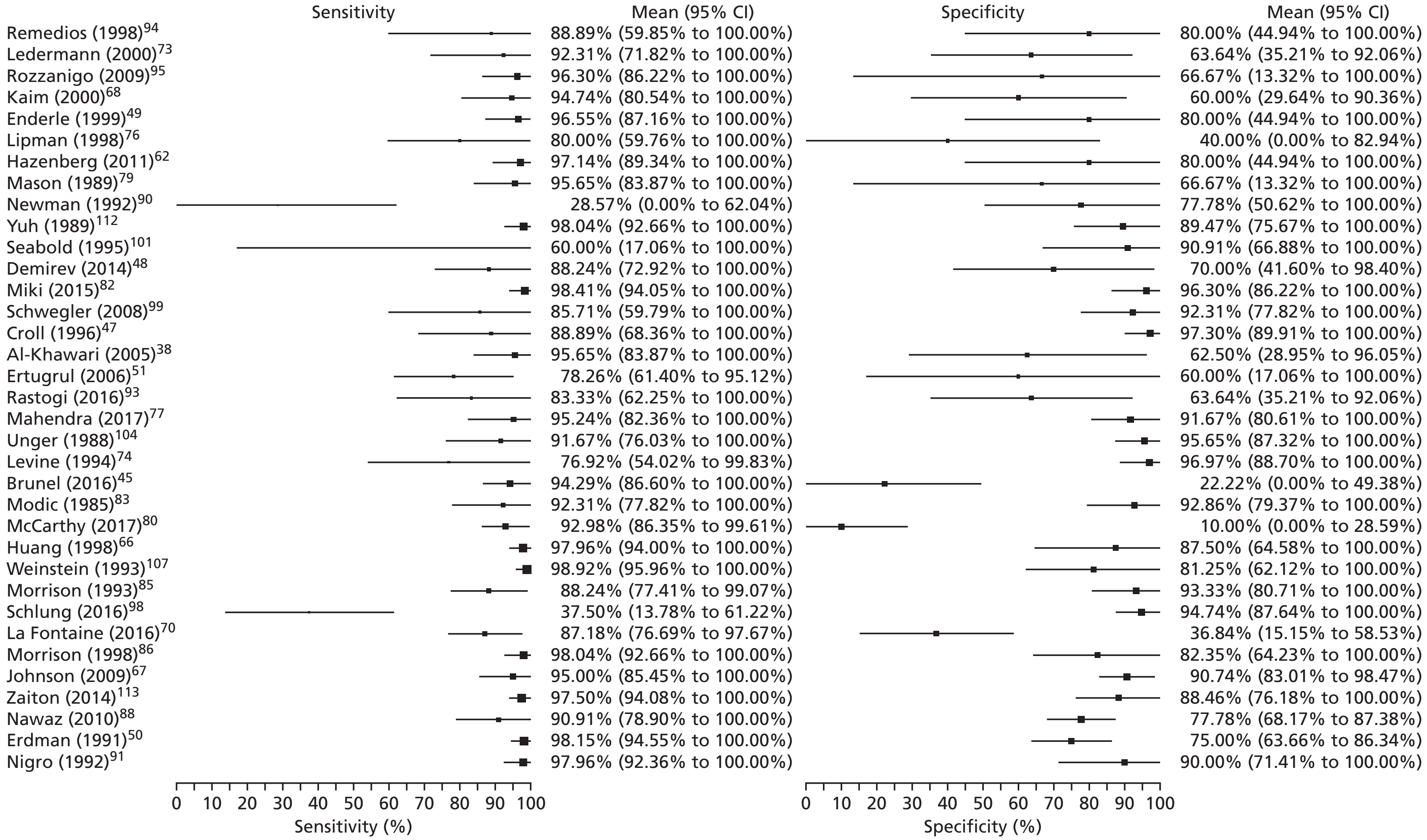
Diagnostic odds ratios are similar for MRI, SPECT and PET, suggesting similar diagnostic accuracy for these tests, as seen in *Figure 4*. The odds ratios are considerably lower for scintigraphy, radiography and CT, suggesting that these tests have poorer diagnostic accuracy. MRI and SPECT have high PRs of > 70%, so most people have an imaging test result that would lead to being diagnosed with osteomyelitis. This, in part, reflects high incidence in the studies. PET, however, has a lower PR of 52.7%. This may be because the higher specificity and lower sensitivity of PET reduce the number of positive PET scans overall, and in particular the number of false-positive diagnoses.

Comparisons between tests

Forty of the included studies of adults reported diagnostic accuracy data for two or more imaging tests. To determine whether or not these studies show evidence that one test is superior to another, we compared the DORs for each test within each study. *Appendix 4*, *Figure 23*, shows these DORs for each imaging test in each study. These results show that, when radiography was included, it was generally inferior to other tests (particularly MRI and scintigraphy); scintigraphy was generally inferior to MRI.

In *Figure 6*, each dot is the difference in log-DOR between the test in the title of each subplot and the comparator test, indicated by the colour of the dot. Points to the right of the central black line are where the ‘title’ test has greater DOR than the ‘comparator’ test. These results suggest that SPECT and PET were generally superior to whatever test they were compared with (but with few studies). MRI was superior to scintigraphy and radiography, but not to PET or SPECT. Scintigraphy was superior to radiography, but not MRI, and radiography was generally inferior to all other tests.

To formally assess whether or not these apparent differences are genuine, regression modelling was used to compare studies. This regressed the diagnostic accuracy data to estimate differences in DORs and specificity between tests across studies (see *Chapter 3, Comparison of imaging tests*). Two models were fitted: one using radiography as the baseline test and one using MRI as the baseline test. The formal results of these models are given in *Appendix 4*, *Table 20*, and a summary is presented in *Table 5*.



	Sensitivity Mean (95% CI)	Specificity Mean (95% CI)
Remedios (1998)[94]	88.89% (59.85% to 100.00%)	80.00% (44.94% to 100.00%)
Ledermann (2000)[73]	92.31% (71.82% to 100.00%)	63.64% (35.21% to 92.06%)
Rozzanigo (2009)[95]	96.30% (86.22% to 100.00%)	66.67% (13.32% to 100.00%)
Kaim (2000)[68]	94.74% (80.54% to 100.00%)	60.00% (29.64% to 90.36%)
Enderle (1999)[49]	96.55% (87.16% to 100.00%)	80.00% (44.94% to 100.00%)
Lipman (1998)[76]	80.00% (59.76% to 100.00%)	40.00% (0.00% to 82.94%)
Hazenberg (2011)[62]	97.14% (89.34% to 100.00%)	80.00% (44.94% to 100.00%)
Mason (1989)[79]	95.65% (83.87% to 100.00%)	66.67% (13.32% to 100.00%)
Newman (1992)[90]	28.57% (0.00% to 62.04%)	77.78% (50.62% to 100.00%)
Yuh (1989)[112]	98.04% (92.66% to 100.00%)	89.47% (75.67% to 100.00%)
Seabold (1995)[101]	60.00% (17.06% to 100.00%)	90.91% (66.88% to 100.00%)
Demirev (2014)[48]	88.24% (72.92% to 100.00%)	70.00% (41.60% to 98.40%)
Miki (2015)[82]	98.41% (94.05% to 100.00%)	96.30% (86.22% to 100.00%)
Schwegler (2008)[99]	85.71% (59.79% to 100.00%)	92.31% (77.82% to 100.00%)
Croll (1996)[47]	88.89% (68.36% to 100.00%)	97.30% (89.91% to 100.00%)
Al-Khawari (2005)[38]	95.65% (83.87% to 100.00%)	62.50% (28.95% to 96.05%)
Ertugrul (2006)[51]	78.26% (61.40% to 95.12%)	60.00% (17.06% to 100.00%)
Rastogi (2016)[93]	83.33% (62.25% to 100.00%)	63.64% (35.21% to 92.06%)
Mahendra (2017)[77]	95.24% (82.36% to 100.00%)	91.67% (80.61% to 100.00%)
Unger (1988)[104]	91.67% (76.03% to 100.00%)	95.65% (87.32% to 100.00%)
Levine (1994)[74]	76.92% (54.02% to 99.83%)	96.97% (88.70% to 100.00%)
Brunel (2016)[45]	94.29% (86.60% to 100.00%)	22.22% (0.00% to 49.38%)
Modic (1985)[83]	92.31% (77.82% to 100.00%)	92.86% (79.37% to 100.00%)
McCarthy (2017)[80]	92.98% (86.35% to 99.61%)	10.00% (0.00% to 28.59%)
Huang (1998)[66]	97.96% (94.00% to 100.00%)	87.50% (64.58% to 100.00%)
Weinstein (1993)[107]	98.92% (95.96% to 100.00%)	81.25% (62.12% to 100.00%)
Morrison (1993)[85]	88.24% (77.41% to 99.07%)	93.33% (80.71% to 100.00%)
Schlung (2016)[98]	37.50% (13.78% to 61.22%)	94.74% (87.64% to 100.00%)
La Fontaine (2016)[70]	87.18% (76.69% to 97.67%)	36.84% (15.15% to 58.53%)
Morrison (1998)[86]	98.04% (92.66% to 100.00%)	82.35% (64.23% to 100.00%)
Johnson (2009)[67]	95.00% (85.45% to 100.00%)	90.74% (83.01% to 98.47%)
Zaiton (2014)[113]	97.50% (94.08% to 100.00%)	88.46% (76.18% to 100.00%)
Nawaz (2010)[88]	90.91% (78.90% to 100.00%)	77.78% (68.17% to 87.38%)
Erdman (1991)[50]	98.15% (94.55% to 100.00%)	75.00% (63.66% to 86.34%)
Nigro (1992)[91]	97.96% (92.36% to 100.00%)	90.00% (71.41% to 100.00%)

FIGURE 5 Forest plots of sensitivity and specificity for MRI.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 4 Results of univariate meta-analyses

Test	Outcome	Estimate (%)	95% CI	I^2
CT	Sensitivity	69.68	40.12 to 88.74	0
	Specificity	90.24	57.64 to 98.43	0
	DOR	10.28	3.86 to 27.39	0
	PR	55.8	39.49 to 78.83	0
MRI	Sensitivity	95.56	92.36 to 97.46	0
	Specificity	80.7	70.78 to 87.84	0
	DOR	44.05	22.76 to 85.23	54
	PR	72.79	66.89 to 79.22	25
PET	Sensitivity	85.11	71.52 to 92.86	0
	Specificity	92.77	83 to 97.12	0
	DOR	38.51	17.79 to 83.38	34
	PR	53.00	44.2 to 63.55	0
Scintigraphy	Sensitivity	83.6	71.83 to 91.07	0
	Specificity	70.58	57.72 to 80.82	0
	DOR	8.3	5.36 to 12.85	17
	PR	72.7	66.1 to 79.96	0
SPECT	Sensitivity	95.06	87.82 to 98.09	0
	Specificity	81.99	61.54 to 92.83	30
	DOR	65.3	18.38 to 232	57
	PR	74.1	65.15 to 84.29	0
Ultrasound	Sensitivity	94.07	50.17 to 99.6	0
	Specificity	50.99	3.16 to 97.08	0
	DOR	13.92	1.94 to 99.74	0
	PR	82.08	61.66 to 100	0
Radiography	Sensitivity	70.35	61.64 to 77.8	0
	Specificity	81.5	69.61 to 89.45	7
	DOR	6.39	3.7 to 11.01	52
	PR	53.13	43.64 to 64.69	52

These results largely confirm what was observed from the main diagnostic analysis in *Diagnostic meta-analysis*. Radiography has the lowest diagnostic accuracy (in terms of DOR) of all the tests. Its specificity is similar to that of CT, MRI and SPECT, but it has lower sensitivity than any of these. It has lower sensitivity and specificity than PET, but has higher specificity than scintigraphy.

Magnetic resonance imaging has similar diagnostic performance to SPECT and PET, but has lower specificity (and consequently higher sensitivity) than PET. MRI has higher DORs than CT, radiography or scintigraphy.

Synthesis of specific imaging tests

This section considers in more detail the diagnostic accuracy of each of the included imaging tests.

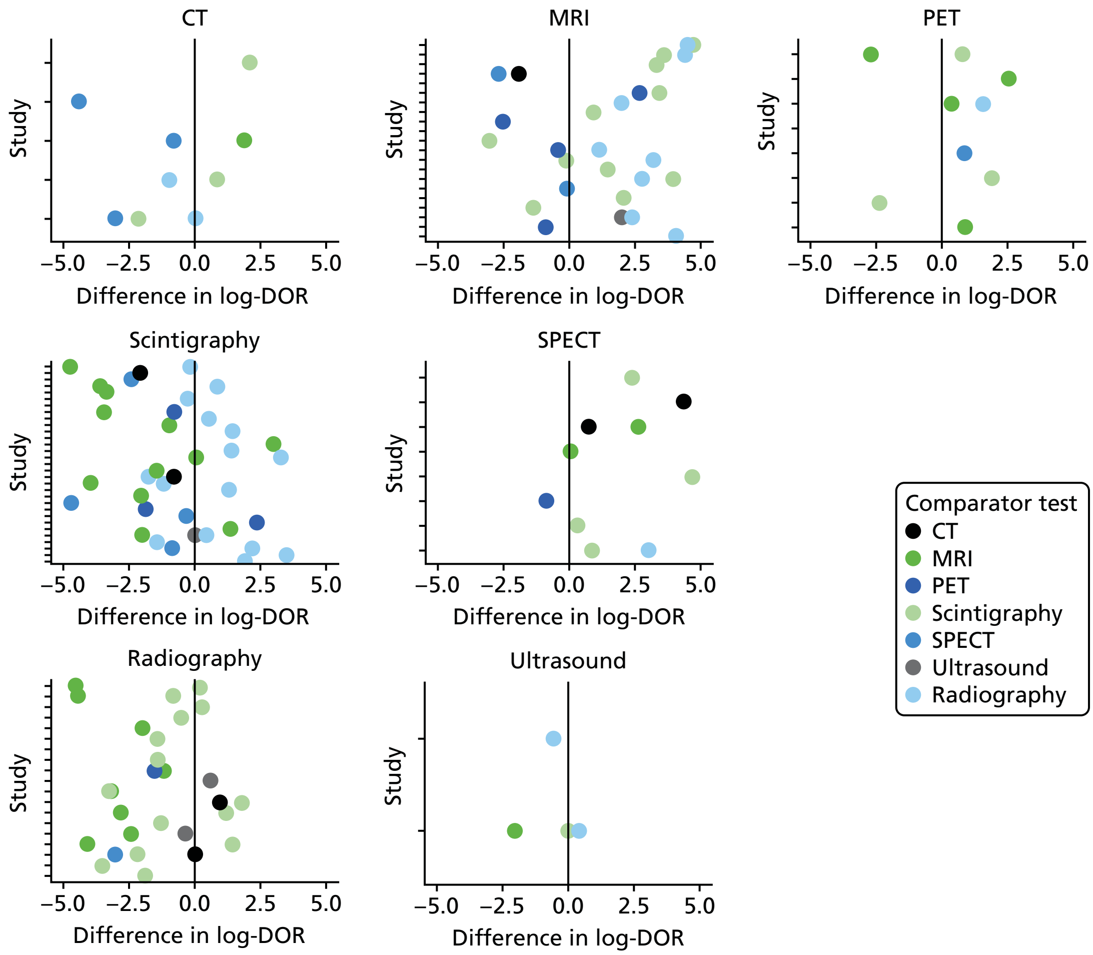


FIGURE 6 Comparing DORs within studies.

TABLE 5 Summary of regression model to compare imaging tests

Parameter	Radiography worse than	Radiography similar to	Radiography better than
DOR	CT, MRI, PET, scintigraphy and SPECT	None	None
Specificity	PET	CT, MRI and SPECT	Scintigraphy
	MRI worse than	**MRI similar to**	**MRI better than**
DOR	None	PET, SPECT	CT, scintigraphy and radiography
Specificity	PET	CT, SPECT and radiography	Scintigraphy

Studies of scintigraphy

Scintigraphy was not specified as an imaging test of interest in our protocol, but was included in many of the eligible studies as a comparator with other tests. Hence, we included it in our main analysis in *Synthesis of diagnostic accuracy in adults*. To facilitate a more complete examination of the diagnostic value of scintigraphy, we identified larger, recent studies (> 20 participants and published since 1990) of scintigraphy for analysis, where scintigraphy was the only imaging test reported in the publication. Nine studies[132,136,142,148–152,165] from our database searches met these criteria and reported sufficient data for meta-analysis. Because they were off-protocol, we note that the additional nine studies have not been formally assessed for risk of bias. These nine studies were combined with the studies of scintigraphy included in the main meta-analysis, in order to explore the impact of this additional evidence on scintigraphy, and as a sensitivity analysis compared with using only the primary included studies.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

A summary of all scintigraphy studies is given in *Appendix 4*, *Table 21*. As with other imaging tests, most of the studies are in people with diabetic foot ulcers. A range of types of scintigraphy have been reported, including (hydroxy)methylene diphosphonate [(H)MDP], WBC, hexamethylpropyleneamine oxime (HMPAO) and gallium scans. Some studies reported combinations of these tests. Three radioisotopes were used: technetium-99m (^{99m}Tc), indium-111 (^{111}In) and gallium-67 (^{67}Ga).

Figure 7 presents a summary ROC plot for the sensitivity and specificity results from these studies, classified by scintigraphy type used. This suggests that there is considerable variation both within and between types of scintigraphy.

A bivariate meta-analysis of all scintigraphy studies gave an overall sensitivity of 85.6% (95% CI 80.9% to 89.2%) and an overall specificity of 71.7% (95% CI 62.1% to 79.6%). This is similar to the result seen in the main analysis (see *Figure 3*), but with a slightly higher sensitivity, suggesting that the additional studies (of scintigraphy only) have higher sensitivity than those where scintigraphy was compared with other tests. This could be as a result of either bias against scintigraphy in comparative studies, or poorer quality of the scintigraphy-only studies.

Figure 8 shows the results of bivariate meta-analyses categorised according to the type of scintigraphy performed. This suggests that ^{99m}Tc HMPAO WBC scintigraphy and mixed ^{111}In WBC with ^{99m}Tc MDP (methylene diphosphonate) scintigraphy can have very high diagnostic accuracy, with sensitivity > 85% and specificity > 90%; this is similar to the diagnostic accuracy of PET, as seen in *Figure 3*. The ^{99m}Tc (H)MDP scintigraphy, by contrast, has a much lower specificity, of < 50%.

Studies of magnetic resonance imaging scans

Most studies reported only one diagnostic accuracy result for MRI, and there was insufficient information to distinguish between different types of MRI scan. One study[86] reported results for four types of MRI, with no evidence of any difference in diagnostic accuracy among them.

The MRI results in the main analysis had a very wide range of specificities, despite fairly consistent sensitivity. We sought to identify any characteristics of the included studies that might explain this variation in specificity.

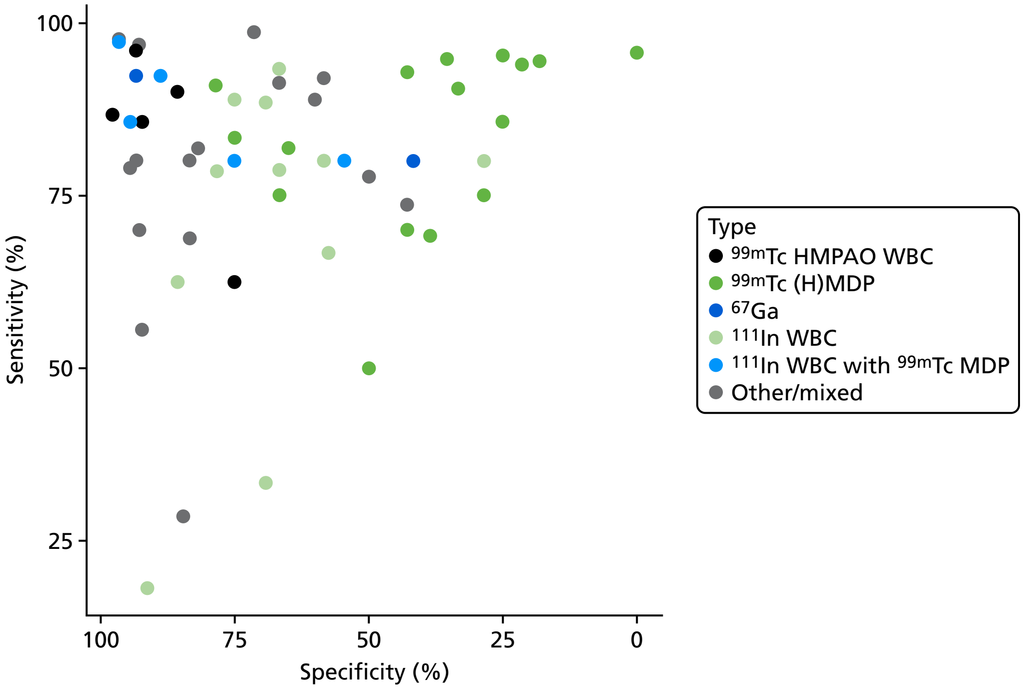


FIGURE 7 Summary ROC plot for scintigraphy studies.

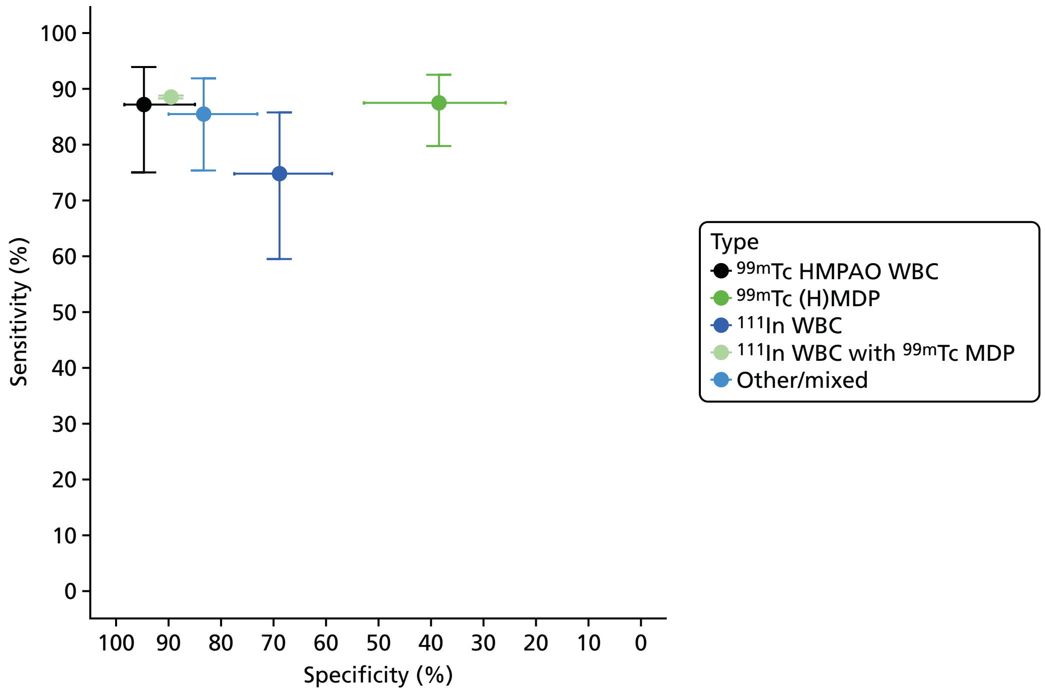


FIGURE 8 Diagnostic accuracy of scintigraphy according to type of test performed.

Appendix 4, *Figure 24*, shows the association between specificity and incidence of osteomyelitis for each study. This shows that specificity declines as incidence increases. This association was statistically significant (a decline in log-odds of specificity of 0.31 per 10% increase in incidence, $p = 0.008$). This suggests that osteomyelitis may be over-diagnosed using MRI in populations in which osteomyelitis is common. The model was not a good fit (adjusted $R^2 = 0.17$), so incidence does not explain all the variation in specificity. No other possible causes of this variation could be identified.

Studies of single-photon emission computed tomography and single-photon emission computed tomography/computed tomography

As with scintigraphy studies, SPECT studies used a range of radioisotopes and test types to perform SPECT imaging. *Figure 9* shows the ROC plot of sensitivity and specificity according to the type of test used. There are too few studies to distinguish between tests, or to perform a bivariate meta-analysis. The results show no clear evidence of difference between the types of test. When comparing results with those for scintigraphy (see *Figure 8*), the results suggest that SPECT produces higher sensitivity than planar scintigraphy for the same test type.

There was no evidence that studies of SPECT/CT had differing diagnostic accuracy to those of SPECT alone, although the number of studies was limited (results not shown).

Studies of positron emission tomography and positron emission tomography/computed tomography

There was no evidence that studies describing the test used as PET/CT differed in diagnostic accuracy from those describing the test only as PET (see *Appendix 4*, *Figure 25*). All but one study used ^{18}F-FDG PET.

There was no evidence that quantitative analysis (such as measuring the standardised uptake value) improved the accuracy of PET (results not shown).

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

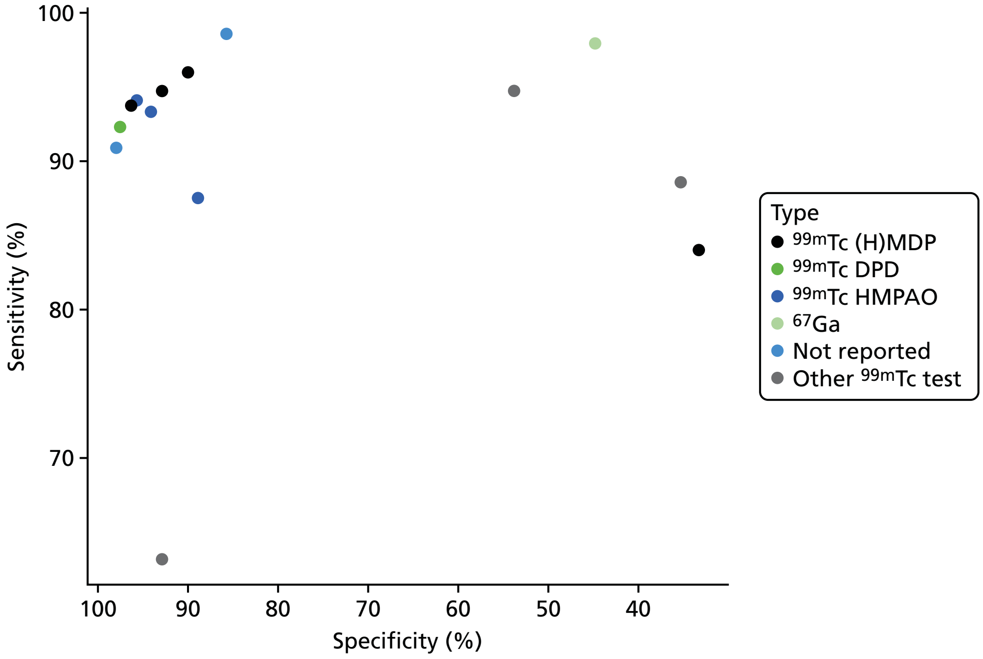


FIGURE 9 Diagnostic accuracy of SPECT studies according to type of SPECT used.

Studies of ultrasound

Only two studies of ultrasound for diagnosing osteomyelitis in adults were identified, so they could not be included in any bivariate meta-analysis. The results of these studies are summarised in *Table 6*. The two studies are not consistent in their estimated sensitivity and specificity. Both have DORs that suggest poorer diagnostic performance than MRI, PET or SPECT (see *Table 4*).

Studies reporting on combinations of imaging tests

Twelve studies[40,43,54,55,64,65,69,75,84,96,109,141] reported data on the diagnostic accuracy of combining two tests. These were mostly combinations of scintigraphy with SPECT. Two studies combined radiography with probe-to-bone tests. These are summarised here, although we note that probe to bone is not an imaging test and so is not covered in this review.

Figure 10 shows the sensitivity and specificity estimates for all the combination tests. This shows considerable variation in the diagnostic accuracy of combination tests. Most tests have too few studies to draw any conclusions. A bivariate meta-analysis of the SPECT + scintigraphy studies gives a summary sensitivity of 87.3% (95% CI 87.2% to 87.5%) and specificity of 83.4% (95% CI 83.3% to 83.6%). This is a lower sensitivity and specificity than was estimated for SPECT alone (see *Table 4*).

TABLE 6 Summary of diagnostic accuracy of ultrasound studies

	Study	
	Enderle *et al.* (1999)[49] (*n* = 19)	**Nath and Sethu (1992)[87] (*n* = 25)**
	Cause: diabetic foot ulcers	**Cause: long bone pain or swelling**
Sensitivity (%) (95% CI)	78.6 (57.1 to 100)	96.8 (88 to 100)
Specificity (%) (95% CI)	80 (44.9 to 100)	30 (1.6 to 58.4)
DOR (95% CI)	14.7 (1.2 to 185.2)	12.9 (0.6 to 292.8)
PPV (%) (95% CI)	91.7 (76 to 100)	68.2 (48.7 to 87.6)
NPV (%) (95% CI)	57.1 (20.5 to 93.8)	85.7 (49.1 to 122.4)

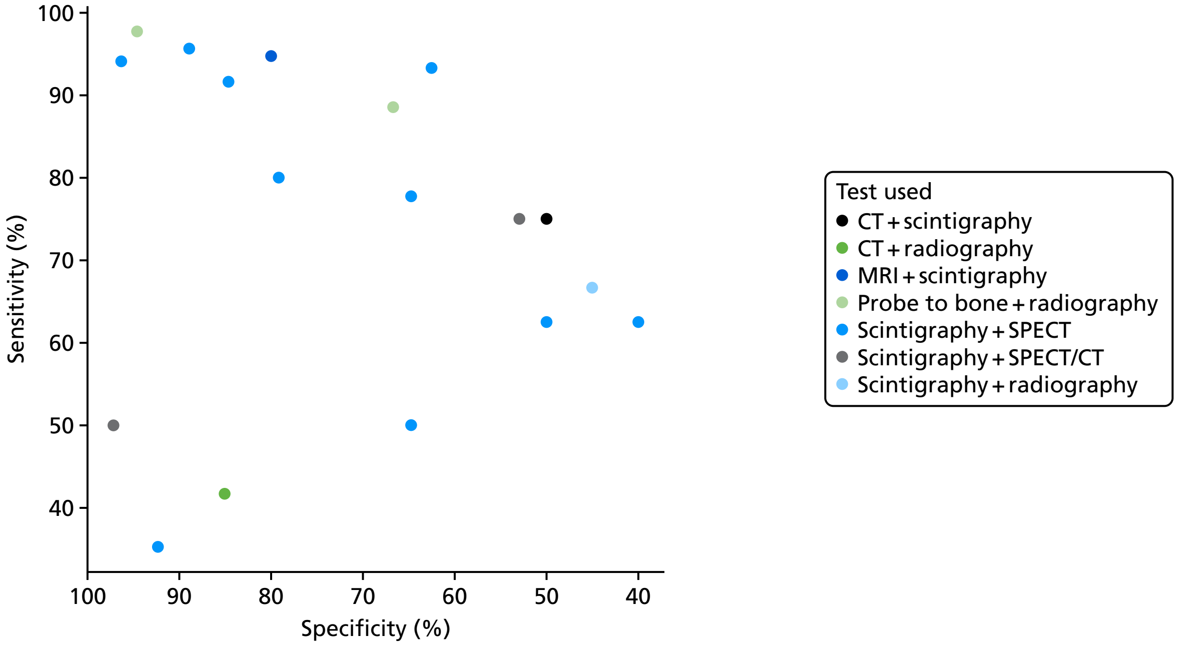


FIGURE 10 Diagnostic accuracy results for combination tests.

Figure 11 compares the DORs for the combination tests with those for the individual tests within each study. This shows no evidence that combining tests improves diagnostic accuracy; in most studies the combined test (circle) has a DOR no higher, and often lower, than the best single test (triangle).

One study (Aragón-Sánchez *et al.*[40]) of radiography combined with probe to bone obtained very high diagnostic accuracy (sensitivity 97.7%, specificity 94.6%). However, this study also achieved much higher diagnostic accuracy for radiography alone than was observed in most other studies. It is unclear whether this result is meaningful or a chance result in one study.

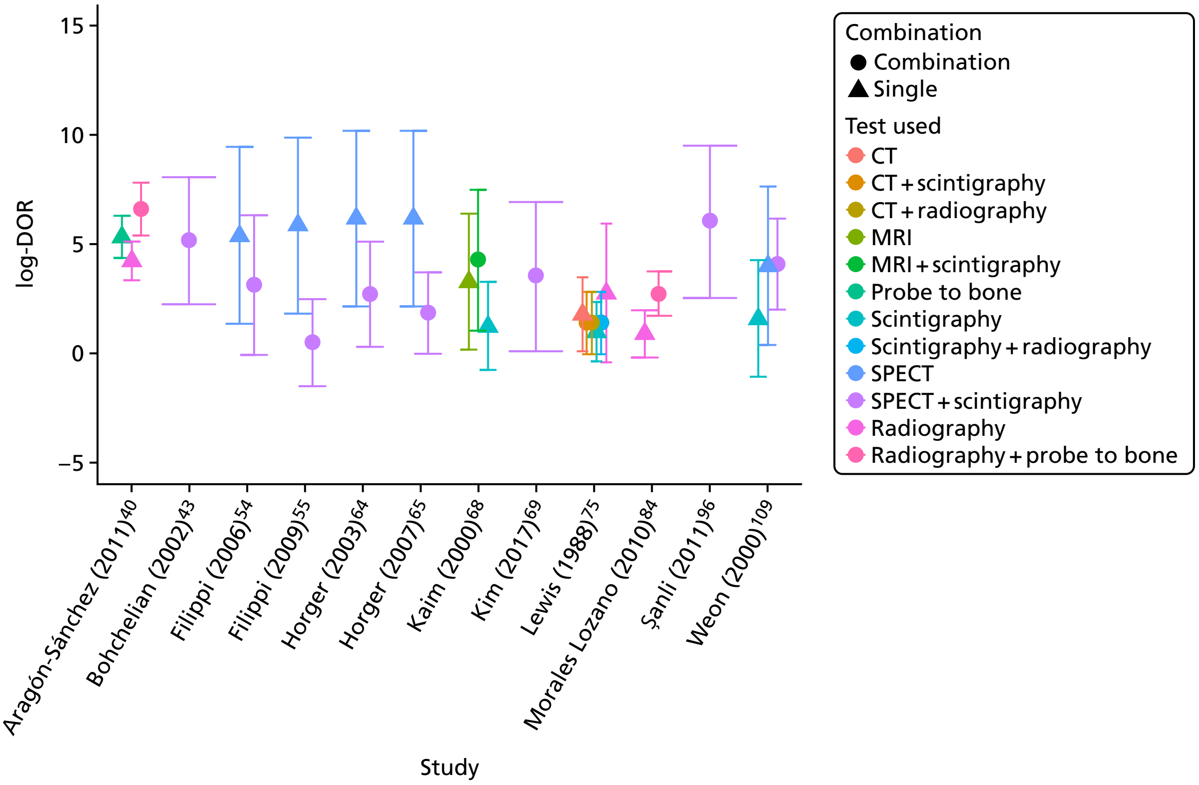


FIGURE 11 Diagnostic odds ratios for combination tests.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Other factors and subgroups

This section considers other factors that may affect diagnostic accuracy: the choice of reference standard, risk of bias, the presence of indwelling metalwork, the use of imaging tests prior to the main test and whether osteomyelitis is acute or chronic.

Choice of reference standard

The eligible reference standards for confirming the absence of osteomyelitis were histopathology, microbiology or clinical follow-up of at least 6 months. It is possible that the choice of reference standard could affect the diagnostic accuracy.

In *Appendix 4*, *Figure 26* shows the ROC plots of diagnostic accuracy in accordance with the reference standard, for each imaging test. These results suggest that, when only clinical follow-up was used, specificity might be higher than average, particularly for MRI. A formal regression model to assess this, however, found no statistically significant evidence of a difference between reference standards in either specificity or DOR. We conclude that choice of reference standard is unlikely to lead to bias in the main meta-analyses.

Study quality and risk of bias

The impact of the QUADAS-2 assessment of each study on diagnostic accuracy was investigated. In *Appendix 4*, *Figure 27* shows the ROC plots of sensitivity and specificity for each study according to whether the study was rated as being at a high, unclear or low risk of bias for patient selection. In *Appendix 4*, *Figure 28* repeats this for quality of the index test.

Neither figure suggests that there is any evidence that diagnostic accuracy results were biased by quality factors. Reference standard and patient flow issues were rated as being at a low risk of bias for most studies, and there was no evidence that these had any impact on diagnostic accuracy (results not shown).

Indwelling metalwork

Few studies reported on the number of patients (if any) who had indwelling metalwork at the time of their scan. In *Appendix 4*, *Figure 29* shows the diagnostic accuracy data for each imaging test according to whether indwelling metalwork was present in some patients, not present in any, or unreported. This figure shows no evidence that indwelling metalwork alters diagnostic accuracy.

Prior use of other imaging tests

In some studies, patients were included only after performing a prior imaging test, usually radiography. Patients would go on to the main imaging test only if osteomyelitis had not been ruled out by the initial test. This extra testing could affect the diagnostic accuracy observed.

Figure 12 shows the sensitivity and specificity for all imaging tests, according to the prior imaging test used (if any). This suggests that prior use of radiography might improve specificity in MRI studies, and both sensitivity and specificity in scintigraphy studies.

Formal regression analysis to test this possibility found no statistically significant evidence that radiography prior to MRI improved specificity ($p = 0.472$) or DOR ($p = 0.201$). There was some evidence (but not conventionally statistically significant) that radiography prior to scintigraphy increased the DOR ($p = 0.141$). Hence it is possible, but uncertain, that using radiography as a precursor test to MRI or scintigraphy may improve diagnostic performance.

Other than radiography, few tests were used in addition to the main test. This analysis considered only instances when prior tests were explicitly reported. In practice, it is likely that radiography will be used as an initial test in most patients.

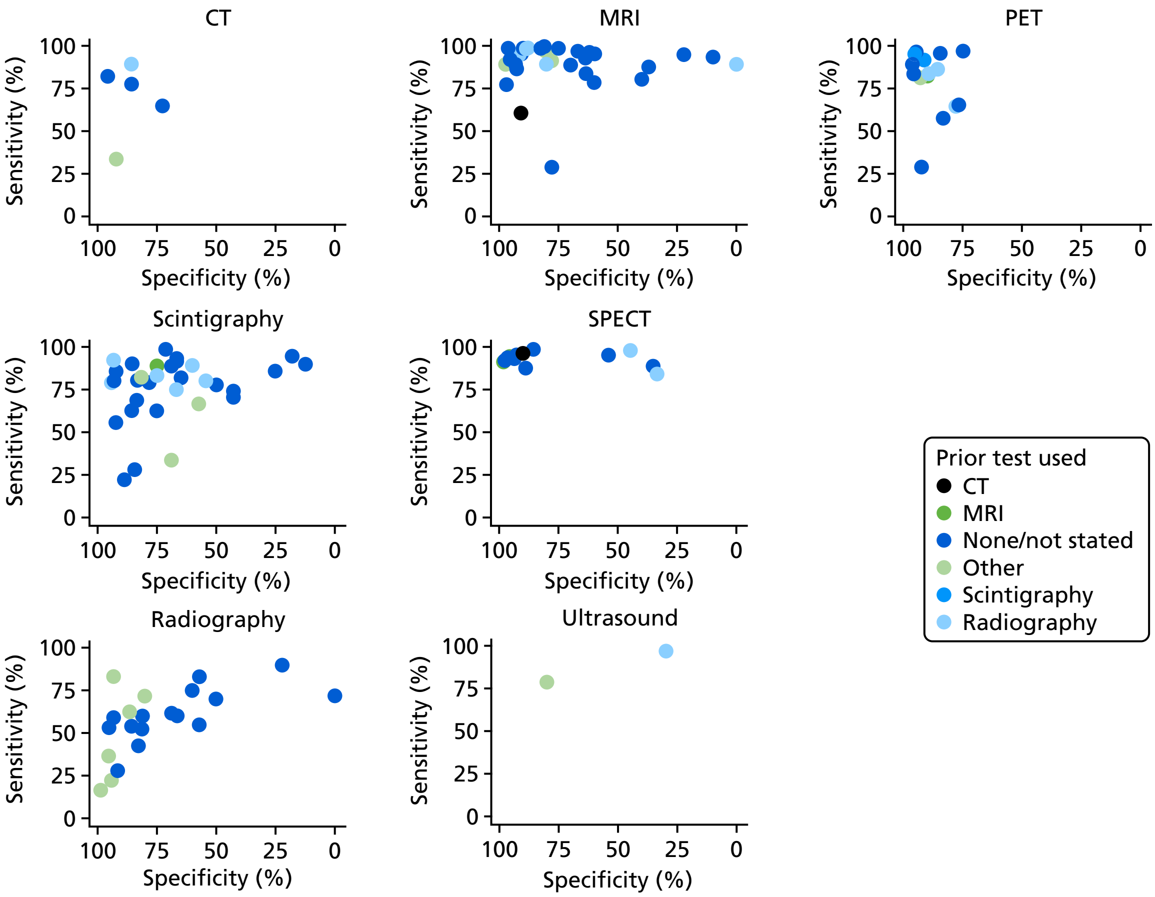


FIGURE 12 Sensitivity and specificity according to whether or not any imaging test was used before the main test.

Acute and chronic osteomyelitis

Studies were divided according to whether the suspected osteomyelitis was acute or chronic. Studies were categorised as:

- acute or subacute osteomyelitis
- COM
- COM and AOM
- not reported.

Figure 13 shows the sensitivity and specificity results according to the test used and acute or chronic status. There appears to be some pattern in the choice of test used. PET and SPECT have not been evaluated in people with AOM; almost all studies of PET are in people with COM. This may be a consequence of the types of people in these studies; nearly all studies of AOM were in people with diabetic foot ulcers. From these data, there is no clear evidence that diagnostic accuracy for any of the tests differ between COM and AOM.

Figure 14 shows the results of bivariate meta-analysis according to acute or chronic status. There is no compelling evidence that diagnostic accuracy for any test varies with acute or chronic status. It is possible that MRI is more specific (91%) but less sensitive (89%) in acute cases than in chronic cases (78% and 98%, respectively). The results are consistent with those in the overall analysis (see *Figure 3*), with MRI, PET and SPECT generally having greater diagnostic accuracy than radiography, CT or scintigraphy.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

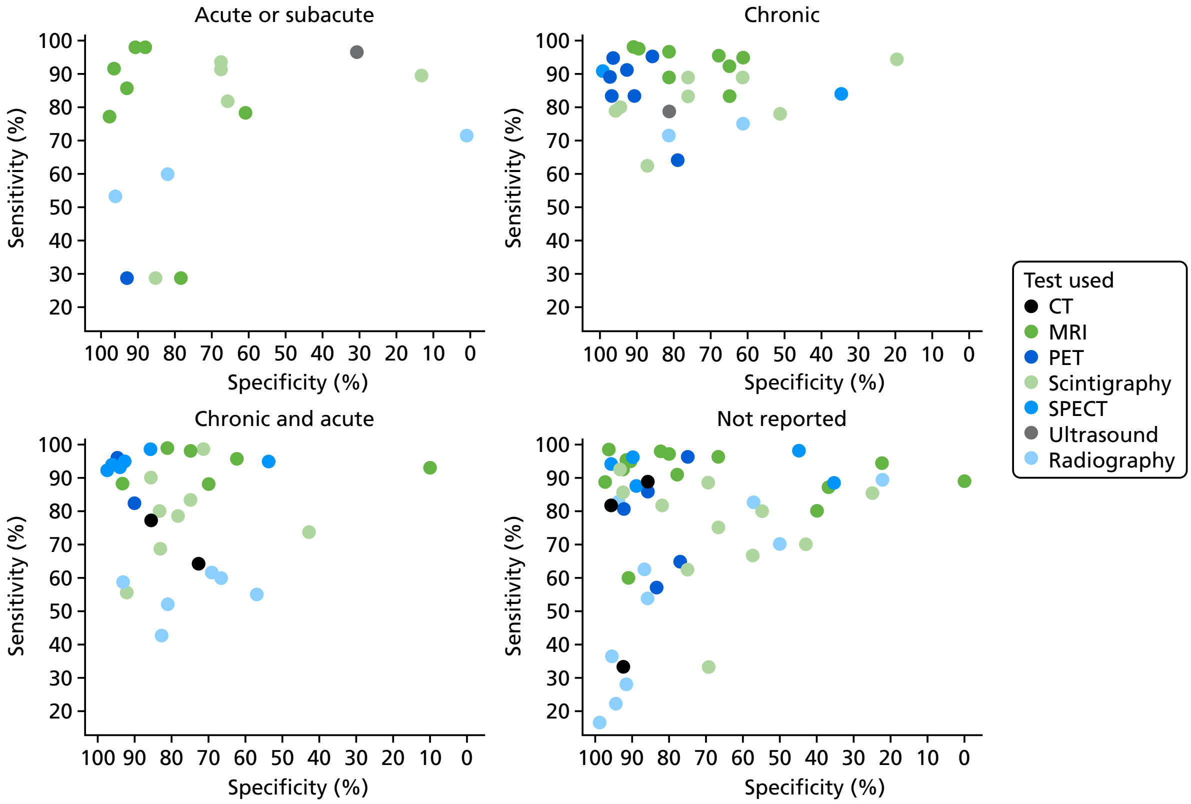


FIGURE 13 Sensitivity and specificity according to whether the study was of AOM or COM.

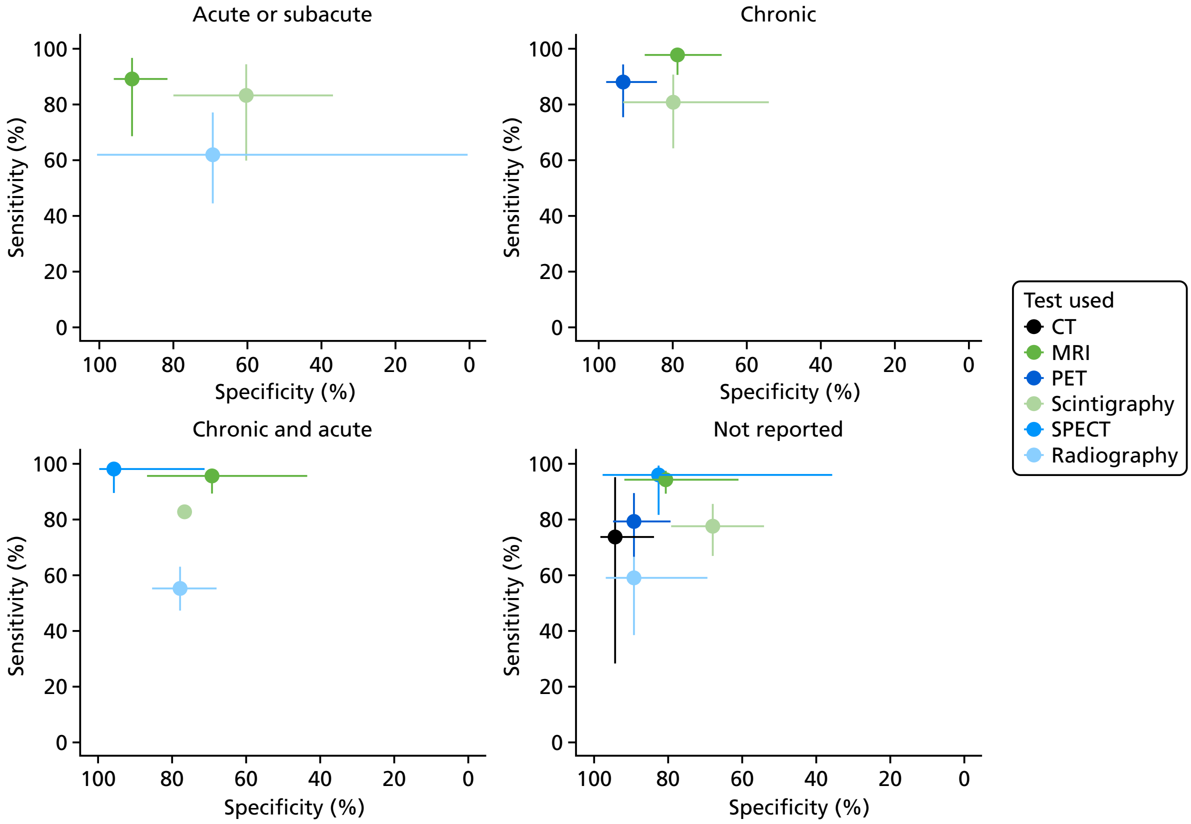


FIGURE 14 Bivariate meta-analysis according to acute or chronic status.

Synthesis of diagnostic accuracy in people with diabetic foot ulcers

This section considers the 35 studies in which diabetic foot ulcers were the primary reason for testing for osteomyelitis.[37,38,40–43,47,49,51,53,55,56,62,67,70,71,74,76,77,82,84,86,88–90,92–95,97,99,102,107,111–113] This includes studies in which at least 60% of patients had diabetic foot ulcers; studies with smaller numbers of people with foot ulcers were excluded. Studies primarily of patients with diabetic foot ulcers constitute around half of the total studies in this review. One study did not report sufficient data to be included in this analysis and is presented in *Studies not included in the quantitative synthesis*.[37]

The sensitivity and specificity estimates from these studies are presented in *Figure 15*. The results suggest high sensitivity (generally > 80%) for MRI and scintigraphy, but with a wide range of specificities. PET showed high specificity, but a range of sensitivities. Radiography generally had low sensitivity. There were too few studies of SPECT or ultrasound to draw any conclusions, and none of CT. In *Appendix 4*, *Figure 30* shows the results for PPV and NPV, with considerable diversity in PPV and NPV values across studies.

Figure 16 shows the results of a bivariate meta-analysis of sensitivity and specificity. The results here are similar to those for all studies (see *Figure 3*). Owing to non-convergence of their respective bivariate analyses, the univariate results are presented for PET and SPECT. Both MRI and PET have high accuracy, with MRI having higher sensitivity but lower specificity. Scintigraphy and radiography have lower specificity, but here scintigraphy has higher sensitivity than PET. The summary HSROC curves (see *Appendix 4*, *Figure 31*) suggest that MRI and PET (and possibly SPECT, although data are limited) have similar overall HSROC curves, and so similar diagnostic accuracy, whereas radiography and scintigraphy have lower accuracy.

In *Appendix 4*, *Figure 32* gives the results of the bivariate meta-analysis of PPV and NPV, which shows that both MRI and PET have high PPV (88%) and NPV (> 85%), with MRI having a higher NPV (95%). Scintigraphy and radiography have lower PPV and NPV rates.

In *Appendix 4*, *Table 23* gives the results of univariate meta-analyses. The results are consistent with bivariate analyses. As in the overall analyses, PET has a markedly lower PR (46%) than MRI (70%). This means that fewer people are considered to have osteomyelitis using PET; this is a consequence of its higher specificity and lower sensitivity.

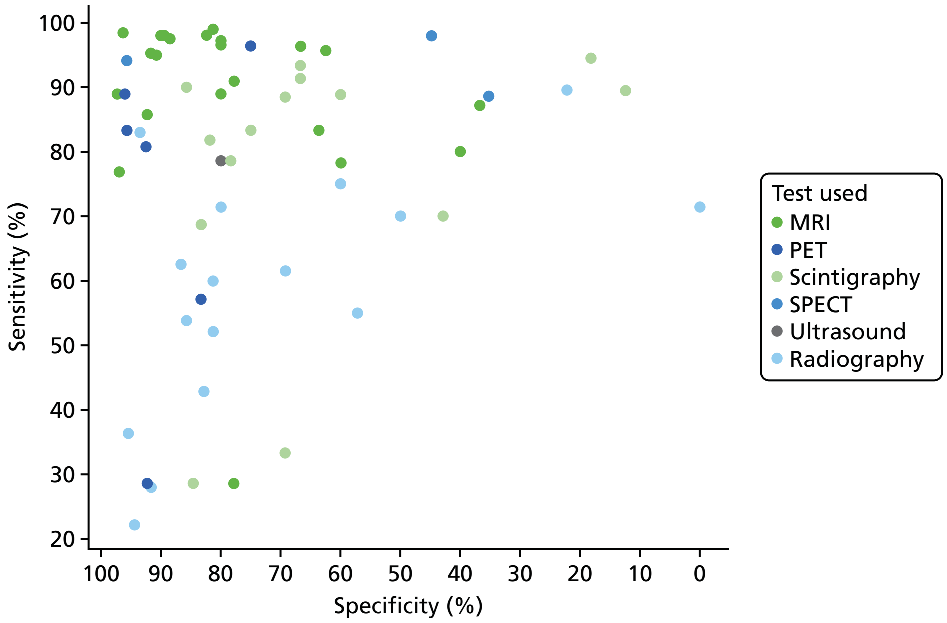


FIGURE 15 Sensitivity and specificity in studies of patients with diabetic foot ulcers.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

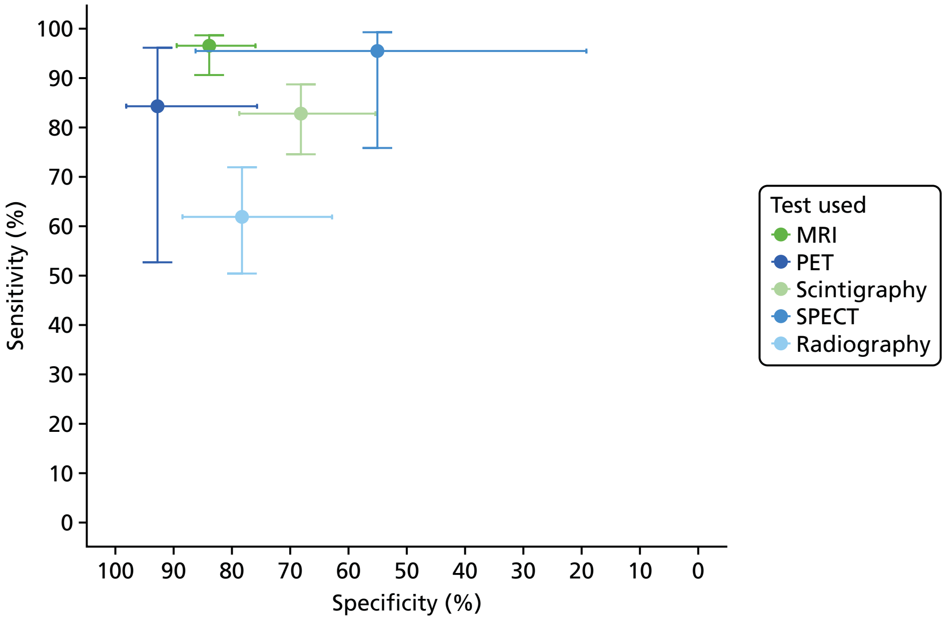


FIGURE 16 Bivariate analysis of sensitivity and specificity in people with diabetic foot ulcers.

In studies using scintigraphy in patients with diabetic foot ulcers, a range of different types of scintigraphy were used. In *Appendix 4*, *Figure 33* shows the results of bivariate meta-analysis of these scintigraphy studies according to the test used. This shows substantial variation in diagnostic accuracy, with ^{99m}Tc HMPAO WBC scintigraphy being notably most accurate (89% sensitivity for 92% specificity). This makes it more sensitive than PET, and more specific, but less sensitive than MRI.

Comparisons of tests

We also considered studies that compared different imaging tests in people with diabetic foot ulcers. In *Appendix 4*, *Figure 34* shows the difference in DORs between tests within studies. This suggests that MRI was generally superior to radiography and scintigraphy, and scintigraphy was moderately superior to radiography. It should be noted that this includes all types of scintigraphy together and, as shown in *Figure 8*, ^{99m}Tc HMPAO WBC scintigraphy may have better diagnostic accuracy. There were too few studies of other comparisons to draw any conclusions. Results from logistic regression models to compare the tests are summarised in *Table 7* (for full results, see *Appendix 4*, *Table 24*). These results are in agreement with the analysis using all studies (see *Table 5*).

TABLE 7 Comparison of tests in people with diabetic foot ulcers

Parameter	Radiography worse than	Radiography similar to	Radiography better than
DOR	MRI, PET and scintigraphy	None	None
Specificity	PET	MRI	Scintigraphy
	MRI worse than	**MRI similar to**	**MRI better than**
DOR	None	PET	Scintigraphy and radiography
Specificity	PET	Radiography	Scintigraphy

Synthesis of diagnostic accuracy for osteomyelitis not related to diabetes

The systematic review identified 31 studies in adults where osteomyelitis was not explicitly related to diabetes.[39,44–46,48,50,54,57–61,63–66,68,73,75,79–81,83,85,87,101,103,105,106,108,109] This excludes studies where some of the participants had diabetes, but includes studies where diabetes was not reported, or where the studies had a range of inclusion criteria, which may have included people with diabetes.

The studies were classified according to the location of the suspected diabetes as follows:

- axial skeleton (skull and spine)
- long bone (leg)
- pelvis, hip or knee
- multiple locations or unreported location.

Studies were also classified according to the suspected cause of osteomyelitis:

- infections and ulcers
- trauma and surgery
- contiguous disease
- multiple causes
- other or unstated cause.

Figure 17 shows the sensitivity and specificity estimates by imaging test and scan location. Similarly, *Figure 18* shows results by suspected cause. There are generally few studies of any imaging test within each category, so it is not possible to draw firm conclusions. The results are broadly similar to those obtained when

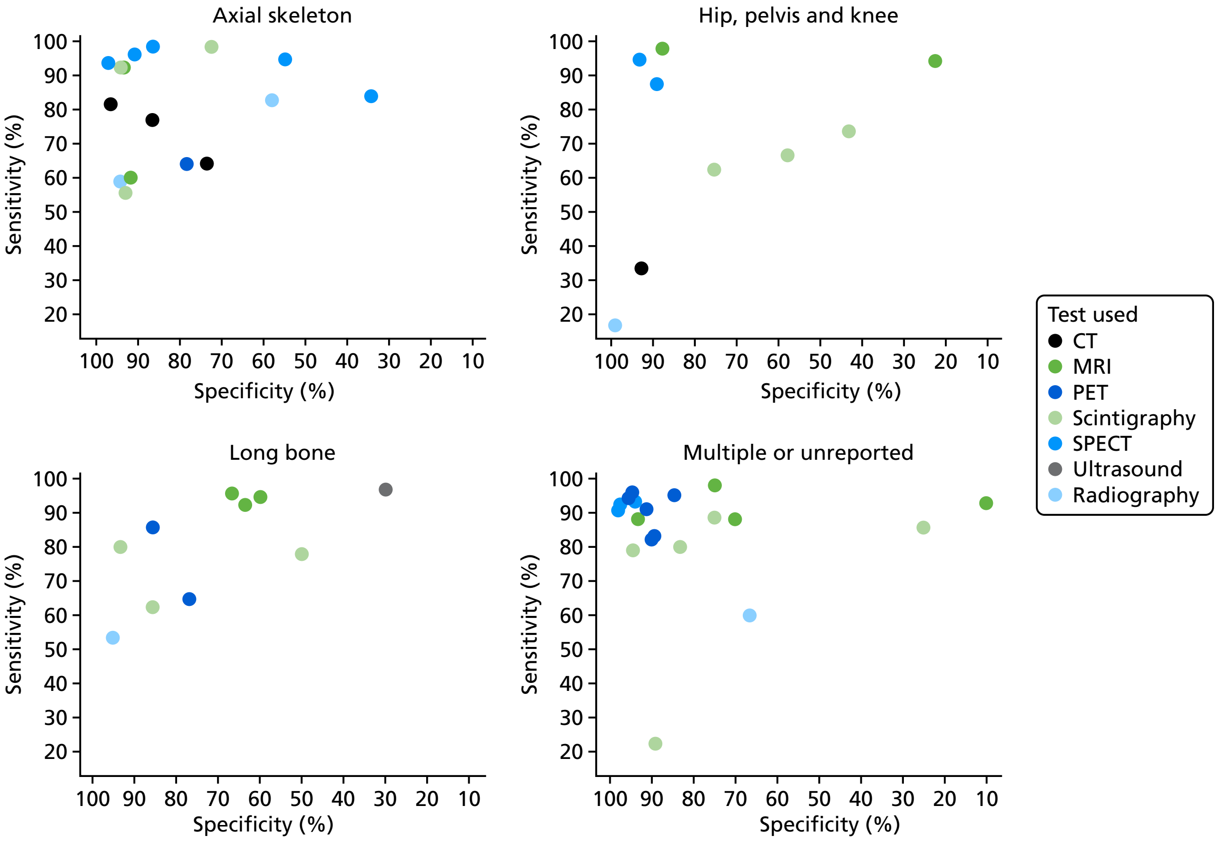


FIGURE 17 Sensitivity and specificity according to scan location.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

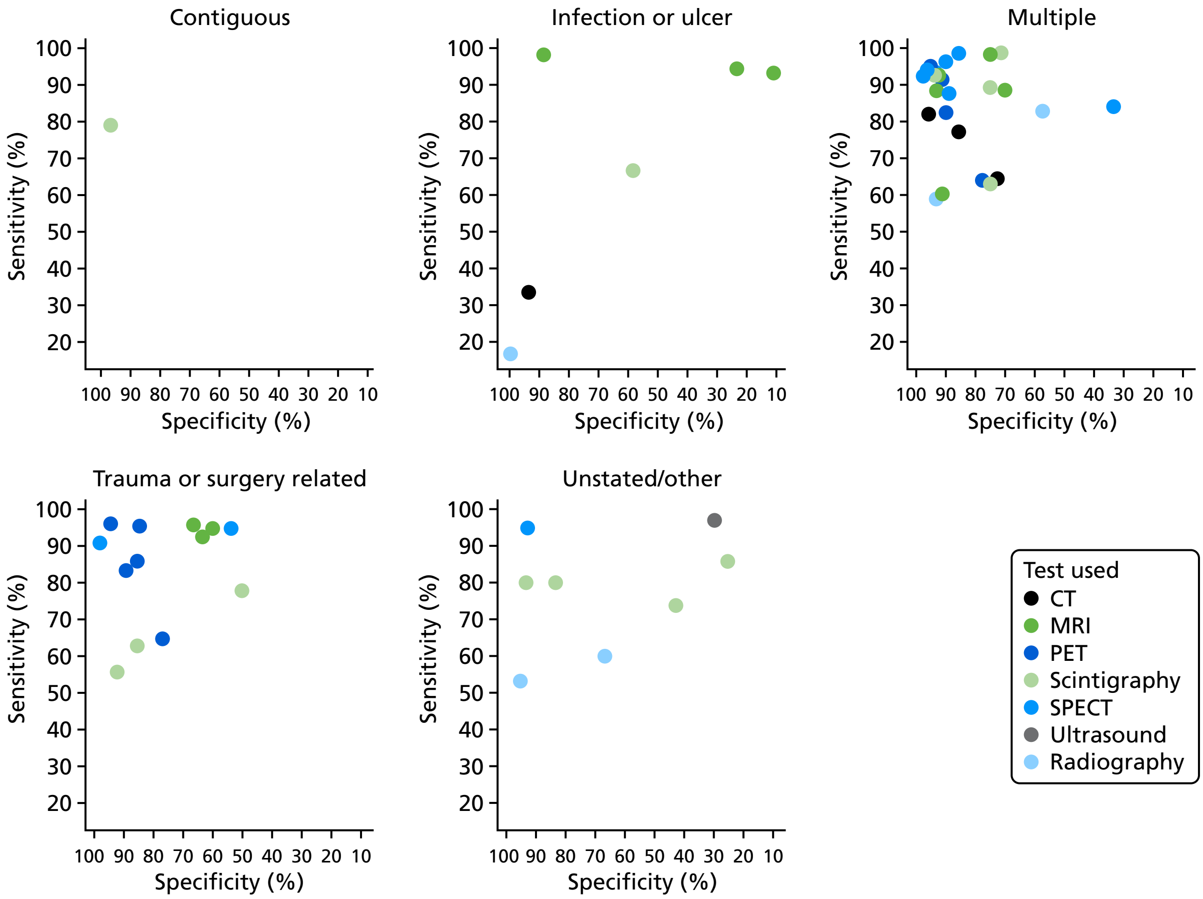


FIGURE 18 Sensitivity and specificity according to likely cause of osteomyelitis.

considering all studies and to the results in studies of diabetic foot ulcers. There is no immediate evidence that diagnostic accuracy varies according to location or cause. The choice of imaging test may vary according to location, with SPECT and CT being used more commonly to image the axial skeleton, where MRI is used less.

Plots for PPV and NPV by location are given in *Appendix 4*, *Figure 35*, and by cause in *Appendix 4*, *Figure 36*. There is no clear evidence that PPV or NPV varies by location of imaging or cause of osteomyelitis.

Given the limited numbers of studies in each location or cause category, bivariate meta-analysis of sensitivity and specificity could be performed for only a few tests. Results of the analyses are shown in *Appendix 4*, *Figure 37*, by location and in *Appendix 4*, *Figure 38*, by cause. The limited number of studies means that CIs for most meta-analysis results are wide, so it is difficult to draw any firm conclusions. There is no clear evidence that diagnostic accuracy varies by location or cause, or that results differ from those seen in the overall meta-analysis of all studies (see *Figure 3*).

Given that there was no clear evidence that diagnostic accuracy for any of the tests varied by cause or location, a combined meta-analysis of all studies in adults without diabetes was performed. The results of the bivariate analysis of sensitivity and specificity are shown in *Figure 19*. Results from univariate meta-analyses are provided in *Appendix 4*, *Table 23*.

These results are broadly comparable to the results using all studies (see *Figure 3*). There is some suggestion that scintigraphy has slightly higher specificity and MRI has slightly lower specificity than the overall results, or to those in diabetic foot ulcer studies (see *Figure 15*), but 95% CIs are too wide to draw any firm conclusions.

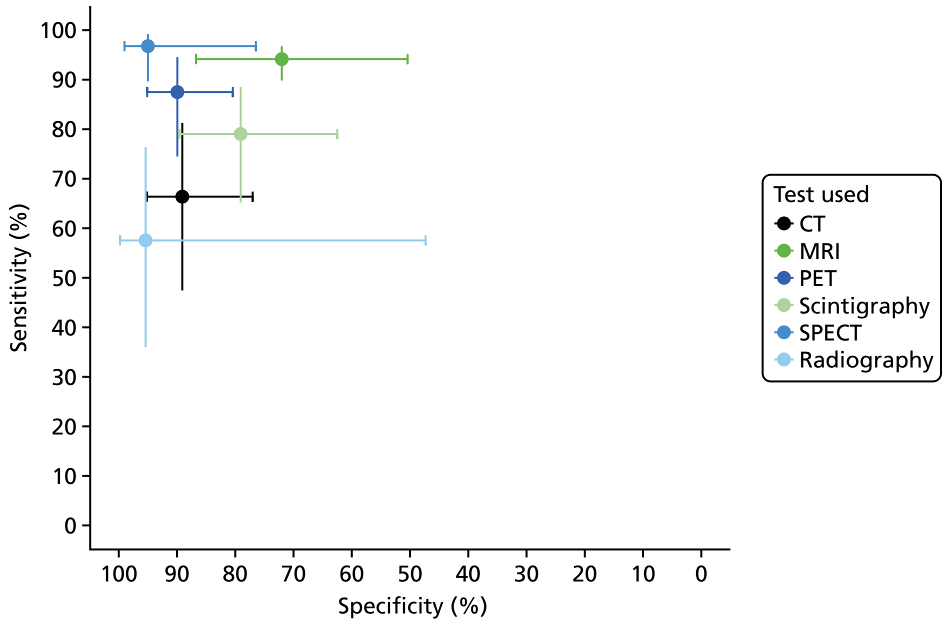


FIGURE 19 Bivariate meta-analysis in studies of people without diabetes.

Comparisons of tests

We also considered studies that compared different imaging tests in people without diabetes. In *Appendix 4*, *Figure 39* shows the difference in DORs between tests within studies (as in *Figure 6* for all studies). This suggests that PET and SPECT are generally superior to radiography and scintigraphy. Results from logistic regression models to compare the tests are summarised in *Table 8* (for full results, see *Appendix 4*, *Table 26*). These appear to be similar to the results for all studies overall, but there were limited data on comparisons between tests.

Studies not included in the quantitative synthesis

Of the two adult studies that were not included in any of the above quantitative syntheses, one evaluated the accuracy of MRI[37] and one compared radiography, CT and radiography + CT[72] (*Table 9*). The results of both studies differed somewhat from the meta-analysis and showed worse diagnostic accuracy overall, although these results should be interpreted with caution owing to the relatively small size of the studies.

Abdel Razek and Samir[37] conducted a prospective study that evaluated MRI in adults with suspected acute diabetic foot osteomyelitis. MRI had a sensitivity of 65% (95% CI 46% to 85%) and a specificity of 61% (95% CI 39% to 84%). These results differ from the meta-analysis, which found higher sensitivity and

TABLE 8 Comparison of tests in people without diabetes

Parameter	Radiography worse than	Radiography similar to	Radiography better than
DOR	MRI, PET and SPECT	CT and scintigraphy	None
Specificity	Too few data to draw conclusions		Scintigraphy
	MRI worse than	**MRI similar to**	**MRI better than**
DOR	None	PET, SPECT	Scintigraphy and radiography
Specificity	Too few data to draw conclusions		

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 9 Studies not included in the quantitative synthesis

First author and year of publication	*n*	Population	Tests	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	DOR
Abdel Razek (2017)[37]	39	Adults with diabetic foot ulcer with suspected AOM	MRI	65 (46 to 85)	61 (39 to 84)	NR
Larson (2011)[72]	44	Adults with pelvic pressure ulcer with suspected pelvis/hip osteomyelitis	Radiography, CT, CT + radiography	Radiography: 88; CT: 50; CT + radiography: 61	Radiography: 32; CT: 85; CT + radiography: 69	NR

NR, not reported.

specificity for MRI (see *Table 4*). Larson *et al.*[72] undertook a retrospective study conducted in patients with suspected osteomyelitis associated with pelvic pressure ulcers. Results for radiography (sensitivity 88%; specificity 32%), CT (sensitivity 50%; specificity 85%) differed from the meta-analysis, which found lower sensitivity and higher specificity for radiography, and higher sensitivity for CT overall (see *Table 4*). The combination of radiography and CT increased the sensitivity of CT alone, but reduced specificity (sensitivity 61%; specificity 69%). This differs from the results of the only other study of radiography + CT reported earlier, which found that radiography and CT combined had a sensitivity of 38% and a specificity of 85% (see *Studies reporting on combinations of imaging tests*).[75]

Synthesis of studies in children

The three studies of children were conducted in single centres in patients with suspected AOM.[52,78,98] Two studies evaluated ultrasonography,[52,78] and two used MRI.[78,98] Radiography, scintigraphy and CT were evaluated in only one study.[78]

The evidence for the accuracy of ultrasound and MRI in children was mixed and limited overall. Ultrasonography had moderate sensitivity (55%, 95% CI 43% to 67%) and specificity (47%, 95% CI 24% to 70%) in 82 children with suspected acute haematogenous osteomyelitis,[78] but had perfect (100%) sensitivity and specificity in a study of 27 children with negative or equivocal initial radiographic findings.[52] MRI had a sensitivity of 81% (95% CI 64% to 93%) and a specificity of 67% (95% CI 22% to 96%) in 38 children with suspected acute haematogenous osteomyelitis,[78] but preoperative MRI had 38% sensitivity and 95% specificity in a study of 54 patients with septic hip. A study-by-study summary is provided in *Table 10*.

Malcius *et al.*[78] conducted a prospective study of 183 hospitalised children with suspected acute haematogenous osteomyelitis. A total of 169 early radiography (median first day of hospital), 142 late radiographic assessments (15th day of hospital stay), 82 ultrasonographies (second day), 76 scintigraphies (third day), 38 MRI scans (seventh day) and 17 CT scans (15th day) were conducted. Scintigraphy (^{99m}Tc-MDP) was conducted in patients who had an unclear diagnosis based on early radiography and/or ultrasound, and MRI and CT were performed when results from scintigraphy were inconclusive. The reference standard included clinical follow-up, microbiology (blood or bone culture) and/or surgery. Acute haematogenous osteomyelitis was confirmed in 156 (85.2%) patients.

Early radiography had low sensitivity (16%, 95% CI 100% to 23%) but high specificity (96%, 95% CI 78% to 100%). Late radiography had high sensitivity (82%, 95% CI 75% to 88%) and specificity (92%, 95% CI 62% to 100%).

Ultrasonography had moderate sensitivity (55%, 95% CI 43% to 67%) and specificity (47%, 95% CI 24% to 70%). Bone scintigraphy had high sensitivity (81%, 95% CI 68% to 90%) and specificity (84%, 95% CI 60% to 97%).

TABLE 10 Diagnostic accuracy results of studies in children

First author and year of publication	Country	Design	*n*	Population	Tests	Reference standard	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	DOR (95% CI)
Malcius (2009)[78]	Lithuania	Prospective	169	Children with suspected acute haematogenous osteomyelitis	Early radiography, late radiography, ultrasound, scintigraphy (^{99m}Tc-MDP), MRI, CT	Physical examination and/or microbiology (blood or bone culture) or surgery	Early radiography: 16 (1 to 23) Late radiography: 82 (75 to 88) Ultrasound: 55 (43 to 67) Scintigraphy: 81 (68 to 90) MRI: 81 (64 to 93) CT: 67 (38 to 88)	Early radiography: 96 (78 to 100) Late radiography: 92 (62 to 100) Ultrasound: 47 (24 to 7) Scintigraphy: 84 (60 to 97) MRI: 67 (22 to 96) CT: 5 (1 to 98)	Early radiography: 4.34 (0.63 to 186.3) Late radiography: 51.17 (6.61 to 2222.0) Ultrasound: 1.08 (0.30 to 3.84) Scintigraphy: 22.3 (4.9 to 132.7) MRI: 8.67 (0.91 to 108.5) CT: 2.0 (0.02 to 172.4)
Schlung (2016)[98]	USA	Retrospective	54	Children with septic hip undergoing preoperative MRI	MRI	Positive joint fluid culture or positive peripheral blood culture with joint aspirate > 50,000 WBCs	37.5 (NR)	94.7 (NR)	NR
Ezzat (2011)[52]	Egypt	Prospective	27	Children with suspected AOM and negative or equivocal plain radiography	Ultrasound (greyscale and power Doppler)	Surgery, cytology, or clinical follow-up	100 (NR)	100 (NR)	NR

NR, not reported.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Magnetic resonance imaging had a sensitivity of 81% (95% CI 64% to 93%) and a specificity of 67% (95% CI 22% to 96%), and CT had a sensitivity of 67% (95% CI 38% to 88%) and a specificity of 50% (95% CI 1% to 98%), although results should be interpreted with caution because of the small number of patients tested.

Overall, Malcius *et al.*[78] found that scintigraphy and MRI are most valuable for detecting acute haematogenous osteomyelitis at the onset of the disease, and late radiography is the most valuable radiological method. The comparability of index tests is limited by the fact that more advanced tests were performed only in harder-to-diagnose subgroups.

Schlung *et al.*[98] conducted a retrospective study of 83 children with suspected septic hip arthritis, of whom 54 underwent preoperative MRI. Osteomyelitis was confirmed by positive joint fluid culture or positive peripheral blood culture with joint aspirate. The sensitivity of MRI was 38% and the specificity was 95%. Of the 10 false-negative cases missed by MRI, six had a positive femoral aspiration that resulted in appropriate management, but three had significant morbidity owing to missed or delayed diagnosis; results for the remaining patient were not reported.

Ezzat *et al.*[52] conducted a small prospective study of children with suspected AOM with negative or equivocal conventional plain radiography. The accuracy of greyscale and power Doppler ultrasound was evaluated in 27 children, of whom 23 underwent surgery. Osteomyelitis was confirmed as positive in 25 out of 27 patients. Ultrasonography had a sensitivity of 100% and a specificity of 100%. The results of this study need to be interpreted with caution because of the limited number of patients.

Review of inter-rater reliability

Characteristics of included studies

Table 11 presents a summary of study and patient characteristics of the 11 studies evaluating the inter-rater reliability of imaging tests for the diagnosis of osteomyelitis.[37,58,59,65,68,86,92,114–117]

Three studies were conducted in Germany,[58,59,65] three in the USA[86,92,116] and two in Spain.[114,115] A single study was conducted in the following countries: Egypt,[37] Switzerland[68] and the UK.[117] Five studies were conducted prospectively[37,58,59,65,114] and six were retrospective.[68,86,92,115–117] One study was conducted in children.[116] The majority of participants were male. The condition targeted was COM in four studies,[58,59,68,117] AOM in one study,[37] AOM and COM in one study,[65,92] and chronic or subacute osteomyelitis in one study.[59,65] Five studies did not report what specific osteomyelitis subtypes were targeted.[86,92,114–116] All studies included experienced raters, although two studies also involved inexperienced radiologists.[114,116]

Five studies included patients with diabetic foot ulcer,[37,86,92,114,115] one study included patients with trauma or surgery as the suspected cause of osteomyelitis,[68] and one included patients with spinal cord injury and pressure ulcers.[117] Three studies did not report the cause of osteomyelitis.[58,59,65]

The inter-rater reliability of the following tests was evaluated: MRI (five studies),[37,68,86,116,117] radiography (two studies),[114,115] PET (two studies),[58,59] scintigraphy (two studies),[58,92] scintigraphy + SPECT (one study),[65] SPECT/CT (one study)[65] and SPECT with CT, MRI or radiography (one study).[65] Four studies evaluated the inter-rater reliability of more than one test.[58,65,86,92]

Critical appraisal

Table 12 presents a summary of the results of the critical appraisal. The studies were considered to be of fair quality overall, although most were small and none provided satisfactory measures of intra-rater reliability (the degree of agreement among repeated administrations of a test by a single rater).

TABLE 11 Characteristics of inter-rater reliability studies

First author and year of publication	Country	Design	*n*	Population	Mean (SD) age [range] (years)	% male	Tests
Abdel Razek (2017)[37]	Egypt	Prospective	41	Adults, DFU, suspected AOM	51 [48–72]	54	MRI
Álvaro-Afonso (2013)[114]	Spain	Prospective	123	Adults, DFU	65 (13) [NR]	72	Radiography
Álvaro-Afonso (2014)[115]	Spain	Retrospective	37	Adults, DFU	58 (14) [NR]	NR	Radiography
Averill (2009)[116]	USA	Retrospective	42	Children, suspected non-spinal osteomyelitis	NR	NR	MRI[a]
Guhlmann (1998)[58]	Germany	Prospective	51	Adults, suspected COM	49 [22–81]	76	PET, scintigraphy[b]
Guhlmann (1998)[59]	Germany	Prospective	31	Adults, suspected COM	48 [20–78]	74	PET
Hauptfleisch (2013)[117]	UK	Retrospective	37	Adults with spinal cord injury, suspected pelvic COM	52 [22–83]	70	MRI
Horger (2007)[65]	Germany	Prospective	31	Adults, suspected AOM and COM	51 [20–89]	52	Scintigraphy + SPECT; SPECT/CT; SPECT with CT, MRI or radiography
Kaim (2000)[68]	Switzerland	Retrospective	18	Adults, suspected post-traumatic COM in long bone	45 [27–65]	72	MRI
Morrison (1998)[86]	USA	Retrospective	68	Adults, DFU	56 [24–85]	NR	MRI[c]
Park (1982)[92]	USA	Retrospective	39	Adults, DFU	NR [NR]	NR	Scintigraphy[d]

AGA, antigranulocyte antibody; DFU, diabetic foot ulcer; NR, not reported; SD, standard deviation; SE, spin echo; SOM, subacute osteomyelitis; STIR, Short-TI Inversion Recovery.

a Contrast enhanced.

b ^{99m}Tc-MDP + AGAb.

c T1-weighted SE; T2-weighted fast SE with fat suppression; fast SE STIR; gadolinium-enhanced T1-weighted fast SE with fat suppression).

d ^{99}Tc-MDP critical appraisal.

Overall, there were no significant concerns about the applicability of the study participants, raters and conduct of imaging test to the review question. Six studies blinded radiographers to relevant patient clinical information;[37,68,86,92,115,116] this may limit the applicability of these studies to clinical practice as such information would typically be available to clinicians, but may also reduce the risk of bias in the interpretation of imaging test results. Three studies did not state whether or not raters were blinded to each other's ratings, which may limit their validity.[65,68,115] There were no significant concerns about loss to follow-up. Only one study provided an estimate of intra-rater reliability, which was below our prespecified acceptability score ($\kappa = 0.8$).[114] All studies used an appropriate measure for calculating reliability (kappa), except one that used percentage agreement only.[92]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 12 Critical appraisal of inter-rater reliability studies

	Question									
First author and year of publication	1	2	3	4	5	6	7	8	9	10
Abdel Razek (2017)[37]	Y	Y	Y	N	Y	Y	Y	Y	N	Y
Álvaro-Afonso (2013)[114]	Y	Y	Y	N	Y	Y	Y	Y	N[a]	Y
Álvaro-Afonso (2014)[115]	Y	Y	Y	Y	Y	Y	UC[b]	Y	N	Y
Averill (2009)[116]	Y	Y	Y	N	Y	Y	Y	Y	N	Y
Guhlmann (1998)[58]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Guhlmann (1998)[59]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Hauptfleisch (2013)[117]	Y	Y	Y	UC[b]	Y	Y	Y	UC[b]	N	Y
Horger (2007)[65]	Y	Y	Y	Y	Y	Y	UC[b]	Y	N	Y
Kaim (2000)[68]	Y	Y	Y	N	Y	Y	UC[b]	Y	N	Y
Morrison (1998)[86]	Y	UC[b]	Y	N	Y	Y	Y	UC[c]	N	Y
Park (1982)[92]	Y	Y	Y	N	Y	Y	Y	Y	N	N

N, no; UC, unclear; Y, yes.

a Experienced 0.75, $p < 001$; moderately experienced 0.61, $p < 001$; inexperienced clinicians 0.57, $p < 001$.

b Insufficient information reported.

c Differing sample sizes for each subtest.

Notes

Reproduced from van de Pol *et al.* (2010).[32] See www.elsevier.com/about/policies/open-access-licenses/elsevier-user-license. Questions: 1. Was a representative sample of participants used? 2. Was a representative sample of raters used? 3. Is replication of the measurement procedure possible? 4. Was clinical information from participants available to raters and comparable to clinical practice? 5. Were participants' characteristics stable during the study? 6. Were raters' characteristics stable during the study? 7. Were raters blinded to each other's results? 8. Can non-random loss to follow-up be ruled out? 9. Was an estimate of intra-rater reliability validly determined and was it above 0.80? 10. Were appropriate measures used for calculating reliability?

Results of inter-rater reliability studies

Table 13 presents the results of the studies of inter-rater reliability. Overall, the tests that showed the best inter-rater reliability were PET ($\kappa = 0.93$ and 0.96, two studies)[58,59] and scintigraphy ($\kappa = 0.91$ and 95% agreement, two studies).[58,92] The inter-rater reliability of MRI ranged from substantial to near perfect ($\kappa = 0.71$ to 0.93, five studies).[37,68,86,116,117] The inter-rater reliability of radiography ranged from fair to substantial ($\kappa = 0.35$ to 0.40 in one study and $\kappa = 0.77$ in one study).[114,115] One study found near-perfect agreement between raters for SPECT/CT ($\kappa = 0.86$) and SPECT with CT, MRI or radiography ($\kappa = 0.87$), and moderate agreement for scintigraphy + SPECT ($\kappa = 0.54$).[65] There was no evidence that results differed according to radiographer experience or patient population.

Review of implementation

One study of implementation was identified by the review.[118]

Govaert *et al.*[118] was a survey of orthopaedic and trauma surgeons, radiologists and nuclear medicine physicians in the Netherlands on preferred imaging strategies for diagnosing post-traumatic osteomyelitis. The 16-question survey included four patient-based clinical cases of fracture-related osteomyelitis. Each case included relevant medical history of the patient combined with a clinical picture of the affected limb and radiographic results. Participants were asked to select which imaging test(s) they considered most appropriate to diagnose or exclude the presence of post-traumatic osteomyelitis.

TABLE 13 Results of inter-rater reliability studies

First author and year of publication	Population	Test	Raters (*n*)	Replicate observations (*n*)	Raters experience	Result	*p*-value
Abdel Razek (2017)[37]	Adults, DFU, suspected AOM	MRI	2	NR	Experienced[a]	$\kappa = 0.93$	NR
Averill (2009)[116]	Children, multiple	MRI	2	42	Inexperienced vs. very experienced[b]	$\kappa = 0.72$ (95% CI 0.52 to 0.92)	NR
Hauptfleisch (2013)[117]	Adults with spinal cord injury, suspected pelvic COM	MRI	2	41	Experienced	$\kappa = 0.92$ (95% CI 0.84 to 1.01); agreement 88%	< 0.0001
Kaim (2000)[68]	Adults, suspected post-traumatic COM in long bone	MRI	2	19	Experienced	$\kappa = 0.88$	NR
Morrison (1998)[86]	Adults, DFU (85%), others (15%)	MRI[c]	2	57 to 71[d]	Experienced	$\kappa = 0.71$ to 0.92[e]	NR
Guhlmann (1998)[59]	Adults, suspected post-traumatic COM	PET	2	31	Experienced	$\kappa = 0.93$	NR
Guhlmann (1998)[58]	Adults, suspected post-traumatic COM[f]	PET; scintigraphy[g]	2	51	Experienced	PET, $\kappa = 0.96$; scintigraphy, $\kappa = 0.91$	NR
Park (1982)[92]	Adults, DFU with suspected AOM and COM	Scintigraphy[h]	3	36	Experienced[i]	95% agreement	NR
Horger (2007)[65]	Adults with suspected AOM and COM, musculoskeletal infection	Scintigraphy + SPECT; SPECT/CT; SPECT with CT, MRI or radiography	2	31	Experienced	Scintigraphy + SPECT, $\kappa = 0.54$; SPECT/CT, $\kappa = 0.86$; SPECT with CT, MRI or radiography, $\kappa = 0.87$	NR
Álvaro-Afonso (2013)[114]	Adults, DFU	Radiography	6	369 (123 × 3)	Inexperienced to very experienced[j]	$\kappa = 0.35$ to 0.40[k]	< 0.001
Álvaro-Afonso (2014)[115]	Adults, DFU	Radiography	2	NR	Experienced	$\kappa = 0.77$	< 0.01

AGA, antigranulocyte antibody; DFU, diabetic foot ulcer; NR, not reported; SE, shot echo; STIR, short-TI intervention recovery.

a 10 and 20 years.

b 20 years vs. limited (paediatric fellow).

c T1-weighted SE; T2-weighted fast SE with fat suppression; fast SE STIR; gadolinium-enhanced T1-weighted fast SE with fat suppression.

d T1-weighted SE: 71; T2-weighted fast SE with fat suppression: 71; fast SE STIR: 42; gadolinium-enhanced T1-weighted fast SE with fat suppression: 57.

e T1-weighted SE: 0.91; T2-weighted fast SE with fat suppression: 0.71; fast SE STIR: 0.72; gadolinium-enhanced T1-weighted fast SE with fat suppression: 0.92.

f ≥ 23% had indwelling work.

g ^{99m}Tc-MDP + AGAb.

h ^{99}Tc-MDP.

i ≥ 7 years.

j Two inexperienced, two moderately experienced, two very experienced.

k 0.35 (very experienced), 0.39 (moderately experienced), 0.40 (inexperienced).

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

There were 346 responders, comprising 153 trauma surgeons, 104 orthopaedic surgeons, 56 nuclear medicine physicians and 33 musculoskeletal radiologists. There was consensus on the usefulness of radiography in patients with a late infection, but not in patients with acute infection. There was a marked difference between trauma surgeons, and orthopaedic surgeons' choice of nuclear medicine imaging for late infection: trauma surgeons preferred PET and orthopaedic surgeons favoured WBC scintigraphy. Similarly, there was a consistent difference between radiologists and nuclear medicine physicians: musculoskeletal radiologists favoured MRI whereas nuclear medicine clinicians preferred PET/CT. CT and three-phase bone imaging for late fracture-related infections were favoured by orthopaedic surgeons and some musculoskeletal radiologists, but not by trauma surgeons or nuclear medicine physicians. Ultrasound-guided biopsy was regarded by all physicians to have some role in patients with early infection but was not popular for late infections. Preference of imaging was influenced by access to equipment, with clinicians selecting imaging methods they had access to in their own hospital.

Summary of previous systematic reviews

The systematic database searches identified 11 previous systematic reviews of the diagnosis of osteomyelitis.[20–30] Most of the studies included in these reviews were also included in this review, although some did not meet our inclusion criteria. The previous reviews are summarised in *Table 14*.

Four of these reviews considered primarily, or only, people with diabetic foot ulcers.[20,22,23,26] Their conclusions varied, depending on the tests included, but MRI, PET and WBC scintigraphy were all suggested as suitable imaging tests.

The Lauri *et al.*[23] review, which is the most recent review of patients with diabetic foot osteomyelitis, concluded that MRI, PET and WBC scintigraphy have similar sensitivity, but MRI has lower specificity. This is similar to the findings of this review.

Three reviews of osteomyelitis in the general population mostly recommended PET and SPECT as having the best diagnostic accuracy,[25,27,28] but that leucocyte scintigraphy could be useful in some cases, such as for COM in the peripheral skeleton,[25] or when PET is unavailable.[28] Only one of these three reviews considered MRI.[25] One review focused on patients with peripheral post-traumatic osteomyelitis and concluded that WBC [or antigranulocyte antibody (AGA)] scintigraphy with SPECT/CT or ^{18}F-FDG-PET/CT has the best diagnostic accuracy in this population.[21] Another review of MRI in patients with pressure ulcers was inconclusive owing to insufficient evidence.[29]

Only two reviews considered osteomyelitis in children. One focused on calcaneal osteomyelitis. The review included only three diagnostic accuracy studies and was inconclusive owing to the limited evidence.[24] One review focused on haematogenous acute and subacute paediatric osteomyelitis.[30] The review included three studies of radiography that showed low sensitivity (16–20%) but high specificity (80–100%). ^{99}Tc bone imaging had variable sensitivity (53–100%) and specificity (50–100%) (three studies). MRI had the highest sensitivity (80–100%) and specificity (70–100%) (five studies), compared with CT (sensitivity 67%, specificity 50%, two studies) and ultrasound (sensitivity 55%, specificity 47%, one study). Only one of the studies included in this review met our inclusion criteria, which was the largest and most comprehensive study in children.[78] Overall, the diagnostic accuracy findings of this review broadly agree with the evidence in adults.

Clinical effectiveness summary and conclusions

Summary of included studies

The review of diagnostic accuracy included 77 studies. Nearly one-quarter of the studies were considered to be at a high risk of bias.

TABLE 14 Summary of previous systematic reviews

First author and year of publication	Studies (*n*)	Osteomyelitis	Index test(s)	Meta-analysis	Main findings	Conclusions
Dartnell (2012)[30]	7	Haematogenous acute and subacute paediatric osteomyelitis	Any	No	Radiography: sensitivity 16–20%, specificity 80–100% ^{99}Tc scintigraphy: sensitivity 53–100%, specificity 50–100% MRI: sensitivity 80–100%, specificity 70–100% CT: sensitivity 67%, specificity 50% Ultrasound: sensitivity 55%, specificity 47%	Radiographs are essential for exclusion of other diagnoses. MRI is the gold-standard imaging technique. If unavailable or a delay is likely, bone and ultrasound scans usually give adequate diagnostic information
Dinh (2008)[20]	9	Diabetic foot osteomyelitis	Any	Yes	Radiography: sensitivity 54%, specificity 68%, DOR 2.84 MRI: sensitivity 90%, specificity 79%, DOR 24.36 Scintigraphy: sensitivity 81%, specificity 28%, DOR 2.10 WBC scintigraphy: sensitivity 74%, specificity 68%, DOR 10.07	MRI is the most accurate imaging test for DFO
Govaert (2017)[21]	10	Peripheral post-traumatic	Any	No	WBC scintigraphy: sensitivity 50–100%, specificity 0–97% PET: sensitivity 83–100%, specificity 51–100% PET/CT: sensitivity 86–94%, specificity 76–100% WBC scintigraphy + SPECT/CT: sensitivity 100%, specificity 89–97%	WBC (or AGA) scintigraphy with SPECT/CT or ^{18}F-FDG/CT have the best diagnostic accuracy for peripheral PTO

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 14 Summary of previous systematic reviews (*continued*)

First author and year of publication	Studies (*n*)	Osteomyelitis	Index test(s)	Meta-analysis	Main findings	Conclusions
Kapoor (2007)[22]	17	Foot osteomyelitis (mainly diabetic)	MRI, scintigraphy and radiography	Yes	MRI: DOR 42.1 (95% CI 14.8 to 119.9) MRI vs. scintigraphy ($n = 7$): DOR 149.9 vs. 3.6 MRI vs. radiography ($n = 9$): DOR 81.5 vs. 3.3 MRI vs. WBC ($n = 3$): DOR 120.3 vs. 3.4	Accuracy of MRI was markedly superior to that of ^{99m}Tc bone scanning, plain radiography and WBC
Lauri (2017)[23]	29	Diabetic foot osteomyelitis	MRI, PET, WBC scintigraphy	Yes	PET: sensitivity 89%, specificity 92%, DOR 95 WBC ^{111}In scintigraphy: sensitivity 92%, specificity 75%, DOR 34 WBC ^{99m}Tc-HMPAO: sensitivity 91%, specificity 92%, DOR 118 MRI: sensitivity 93%, specificity 75%, DOR 37	MRI, PET and WBC have similar sensitivities, PET and WBC have highest specificities
Mooney (2017)[24]	3	Children (calcaneal osteomyelitis)	MRI and radiography	No	MRI: accuracy 100% Radiography: accuracy 14–71.4%	No clear approach to diagnosis of calcaneal osteomyelitis in children

First author and year of publication	Studies (*n*)	Osteomyelitis	Index test(s)	Meta-analysis	Main findings	Conclusions
Termaat (2005)[25]	23	General (chronic)	Any	Yes	PET: sensitivity 96% (95% CI 88% to 99%), specificity 91% (95% CI 81% to 95%) Bone scintigraphy: sensitivity 82% (95% CI 70% to 89%), specificity 25% (95% CI 16% to 36%) Leucocyte scintigraphy: sensitivity 61% (95% CI 43% to 76%), specificity 77% (95% CI 63% to 87%) Bone + WBC scintigraphy: sensitivity 78% (95% CI 72% to 83%), specificity 84% (95% CI 75% to 90%) MRI: sensitivity 84% (95% CI 69% to 92%), specificity 60% (95% CI 38% to 78%)	PET has highest accuracy. Leucocyte scintigraphy has an appropriate accuracy in the peripheral skeleton, but ^{18}F-FDG-PET is superior in the axial skeleton
Treglia (2013)[26]	9	Diabetic foot osteomyelitis	PET	Yes	PET: sensitivity 74% (95% CI 60% to 85%), specificity 91% (95% CI 85% to 96%), DOR 16.96 (95% CI 2.1 to 139.7), AUC 0.874	^{18}F-FDG-PET and PET/CT showed high specificity and are potentially useful if combined with other imaging methods such as MRI, although data are limited
van der Bruggen (2010)[27]	11	General	PET	No	PET: sensitivity 94–100%, specificity 87–100%	^{18}F-FDG-PET is also useful for diagnosis of osteomyelitis with improved spatial resolution over SPECT imaging, allowing more accurate localisation. Localisation can be further improved by adding CT

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 14 Summary of previous systematic reviews (*continued*)

First author and year of publication	Studies (*n*)	Osteomyelitis	Index test(s)	Meta-analysis	Main findings	Conclusions
Wang (2011)[28]	23	General	PET, scintigraphy	Yes	PET: sensitivity 92%, specificity 92%, AUC 0.97 Scintigraphy: sensitivity 83%, specificity 45%, AUC 0.65 Leucocyte scintigraphy: sensitivity 74%, specificity 88%, AUC 0.91 MOAB scintigraphy: sensitivity 88%, specificity 71%, AUC 0.89	PET is superior to scintigraphy, but leucocyte scintigraphy is reasonable if PET is unavailable
Xu (2017)[29]	3	Pressure ulcer	MRI	No	Sensitivity 98% and 94.3%, specificity 89% and 22.2%. NR for 3rd study	Insufficient evidence

AGA, antigranulocyte antibody; AUC, area under the curve; DFO, diabetic foot ulcer; MOAB, monoclonal antigranulocyte antibody; NR, not reported; PTO, post-traumatic osteomyelitis.

The review of inter-rater reliability identified 11 studies. The studies were considered of fair quality overall, although most were small and none provided satisfactory measures of intra-rater reliability.

Only one study on the implementation of diagnostic test imaging for osteomyelitis from the Netherlands was included. The applicability of its findings to other health-care systems is uncertain.

General conclusions from synthesis

In adults, MRI, PET and SPECT scans all had summary sensitivity and specificity of between 80% and 95%. Similarly, they had PPVs between 85% and 93%, and NPVs between 85% and 97%. All had similar DORs and summary HSROC curves, suggesting that overall diagnostic accuracy was similar across the three tests. Despite this good performance, MRI and SPECT showed a very high variability in specificity across studies; for MRI, specificity ranged from 0% to 98%. Results for PET were generally more consistent.

Radiography and CT generally had lower sensitivity, and scintigraphy had lower specificity than MRI, PET and SPECT. Radiography, CT and scintigraphy all had substantially lower DORs than MRI, PET or SPECT, suggesting poorer diagnostic performance in general. Formal comparisons of the tests in studies that analysed multiple imaging tests confirmed the above conclusions. These analyses also confirmed that MRI generally has higher sensitivity, but lower specificity, than PET.

There were only two studies of ultrasound, which were not consistent in their findings, so no conclusions on the diagnostic value of ultrasound can be drawn.

The diagnostic accuracy of scintigraphy depended on the type of scintigraphy performed. Although sensitivity was similar across types of scintigraphy (75% to 92%), specificity was poor for ^{99m}Tc-(H)MDP (at 36% to 46%), but high for ^{99m}Tc-HMPAO WBC (94%) and mixed for ^{111}In WBC and ^{99m}Tc-MDP scintigraphy (88%). This suggests that newer forms of scintigraphy have diagnostic accuracy similar to MRI, PET and SPECT. This is perhaps unsurprising as most SPECT studies used ^{99m}Tc-HMPAO WBC imaging.

No other differences in performance within tests were identified. In particular, there was no evidence that SPECT/CT differed from SPECT alone, or that PET/CT and PET alone differed in diagnostic accuracy. In general, there were too few data reported to distinguish between subtypes of test and few studies compared subtypes of test. In particular, we found no evidence on the impact of contrast agents for MRI.

There was some evidence that the specificity of MRI declined as the incidence of osteomyelitis in the study population increased. This might suggest that there is overdiagnosis of osteomyelitis when using MRI, particularly when most people tested will have osteomyelitis, perhaps because of a desire not to miss actual cases.

Few studies described the diagnostic accuracy when combining different tests, such as combining scintigraphy and SPECT. There was no evidence that test combinations improved diagnostic accuracy compared with using a single test (MRI, PET or SPECT). Some studies were of patients who had already received radiography prior to their main imaging test. There was some suggestion that performing radiography before MRI could improve diagnostic accuracy, but this was not conclusive.

Diagnostic accuracy by cause or nature of osteomyelitis

The largest group of people receiving imaging tests for osteomyelitis were people with diabetic foot ulcers, representing nearly half of the included studies. Results were broadly consistent with the overall analysis. MRI (96% sensitivity, 84% specificity) and PET (79% sensitivity, 92% specificity) both had high diagnostic accuracy, with MRI having higher sensitivity but lower specificity. ^{99m}Tc-HMPAO scintigraphy had high accuracy (92% sensitivity, 89% specificity), comparable to that of PET and MRI.

In studies of people without diabetes, a range of locations (including long bones, pelvis, knee, hip, spine and skull) and a range of possible causes of osteomyelitis (including infections, trauma and surgery) were tested.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

The numbers of studies in any category were small so the ability to compare studies was limited. There was, however, no clear evidence that location or cause of osteomyelitis had any impact on the diagnostic accuracy of any imaging test.

When combining all studies of people without diabetes, the results were similar to the overall results and to those in people with diabetic foot ulcers. MRI (94% sensitivity, 72% specificity), PET (88% sensitivity, 90% specificity) and SPECT (96% sensitivity, 95% specificity) all had high diagnostic accuracy, although diagnostic accuracy was poorer for radiography, scintigraphy and CT.

Overall, there was no statistically significant evidence that diagnostic accuracy of any test varies according to the location or possible cause of osteomyelitis. The choice of tests did vary, with MRI most studied in people with diabetic foot ulcers. SPECT, by contrast, has been little studied in people with diabetes.

Studies of children, inter-rater reliability and implementation

Only three studies of children were identified. The evidence for the accuracy of ultrasound and MRI in children was mixed and limited overall. Only one study evaluated radiography, scintigraphy and CT in children.

There was some evidence to suggest that MRI has an acceptable degree of agreement among radiologists. The evidence for the inter-rater reliability of other imaging tests was very limited.

There is also very limited evidence on the implementation of imaging tests for the diagnosis of osteomyelitis. Notably, there is a lack of evidence on patient preferences and on the cost-effectiveness of diagnostic imaging tests and pathways for osteomyelitis.

Chapter 5 Discussion

Statement of principal findings

This systematic review identified 81 studies relating to imaging tests for the diagnosis of osteomyelitis. Seventy-seven of these were studies of diagnostic accuracy, of which 69 reported diagnostic accuracy data sufficient to be included in a quantitative synthesis. Thirty-six diagnostic accuracy studies were in patients with diabetic foot ulcers, and 40 were in patients without diabetes.

Eleven studies evaluated the inter-rater reliability of at least one imaging test and one study provided data on clinician opinions on imaging tests for osteomyelitis.

Overall, most diagnostic accuracy studies were small (80% had < 50 participants) and nearly one-quarter were considered as being at a high risk of bias, although poor reporting means that there is significant uncertainty about the quality of most of the studies. The studies of inter-rater reliability were considered of fair quality overall, although most were small and none provided satisfactory measures of intra-rater reliability.

Only one study from the Netherlands on the implementation of diagnostic test imaging for osteomyelitis was included.[118] The applicability of its findings to other health-care systems is uncertain.

Diagnostic accuracy in adults

The overall meta-analysis of diagnostic accuracy in adults found that MRI (the most widely studied test) can detect osteomyelitis with high accuracy (95.6% sensitivity, 95% CI 92.3% to 97.5%; 80.7% specificity, 95% CI 70.7% to 87.8%). There was, however, considerable variation in the specificity across tests. This may partly be as a result of the overdiagnosis of osteomyelitis, particularly in studies with a high incidence of osteomyelitis.

Positron emission tomography also had high diagnostic accuracy (85.1% sensitivity, 95% CI 71.5% to 92.9%; 92.7% specificity, 95% CI 83.0% to 97.1%), with lower sensitivity, but higher specificity, than MRI. SPECT also had high accuracy (95.0% sensitivity, 95% CI 87.8% to 98.1%; 82.0% specificity, 95% CI 61.5% to 92.8%). There was no evidence that PET alone differed in accuracy from PET/CT, nor that SPECT differed from SPECT/CT. MRI, PET and SPECT all had similar DORs and summary HSROC curves, suggesting that the three imaging tests have broadly similar diagnostic performance.

Scintigraphy (83.6% sensitivity, 95% CI 71.8% to 91.1%; 70.1% specificity, 95% CI 57.7% to 80.8%), CT (69.7% sensitivity, 95% CI 40.1% to 88.7%; 90.2% specificity, 95% CI 57.6% to 98.4%) and radiography (70.3% sensitivity, 95% CI 61.6% to 77.8%; 81.5% specificity, 95% CI 69.6% to 89.4%) all had generally inferior diagnostic accuracy to MRI, PET or SPECT. This was confirmed by their lower DORs and poorer HSROC curves. There was evidence that the diagnostic accuracy of scintigraphy depended on the type of scintigraphy used. The most up-to-date forms of scintigraphy, such as ^{99m}Tc HMPAO WBC scintigraphy (87.3% sensitivity, 95% CI 75.1% to 94.0%; 94.7% specificity, 95% CI 84.9% to 98.3%) had higher diagnostic accuracy, similar to that of PET or MRI. However, its estimated sensitivity was lower than for 3D SPECT imaging.

There were insufficient studies of ultrasound to assess its diagnostic accuracy.

Comparisons and combinations of diagnostic tests

In studies that compared two or more imaging tests, informal and formal comparisons of these tests confirmed the findings discussed above. MRI, PET and SPECT all had similar diagnostic accuracy in terms of

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

DORs. MRI had higher sensitivity but poorer specificity than PET. Scintigraphy (all types combined), radiography and CT all had lower DORs than MRI, PET or SPECT.

In studies that reported the diagnostic accuracy of combinations of imaging tests (e.g. of the combination of scintigraphy and SPECT), there was no clear evidence that combining tests improved diagnostic accuracy over using a single test alone, although data were limited.

Studies were generally not clear about whether or not radiography was performed prior to the main imaging test. It was possible that prior use of radiography could improve diagnostic accuracy of scintigraphy or MRI, but data were limited and results were not statistically significant.

Subgroups of patients

The main patient subgroup was patients in whom osteomyelitis was suspected as a complication of diabetic foot ulcers. This group represented nearly half of all included studies. The results of meta-analyses in these patients were similar to those from the main meta-analysis, except that there were too few studies of SPECT or CT to reliably assess diagnostic accuracy. There was no evidence to suggest that diagnostic accuracy in patients with diabetic foot ulcers differed from the overall analysis.

Studies of patients without diabetes were divided according to scan location (axial skeleton, long bone, etc.) and by potential cause of osteomyelitis (trauma or surgery, infection, etc.). Data within each category were generally limited, but there was no clear evidence that diagnostic accuracy varied by scan location or cause, or that results differed substantially from the main analysis.

When all studies of patients without diabetes were combined, results were generally similar to the overall analyses. There was some suggestion that the specificity of MRI and PET was lower than in the overall analysis, but CIs were too wide for this to be conclusive.

Few studies reported whether osteomyelitis was acute or chronic, but there was no compelling evidence that this affected diagnostic accuracy. Similarly, there was no evidence that the presence of indwelling metalwork or the choice of reference standard test (microbiology or clinical follow-up) altered diagnostic accuracy. There was no evidence of systematic bias in the results; diagnostic accuracy did not appear to vary according to risk of bias as assessed using QUADAS-2.

Diagnostic tests in children

The evidence for the accuracy of ultrasound and MRI in children was mixed and limited overall. Ultrasonography had moderate sensitivity and specificity in children with suspected acute haematogenous osteomyelitis but had perfect sensitivity and specificity in a small study of children with negative or equivocal initial radiographic findings. MRI had good sensitivity and specificity in children with suspected acute haematogenous osteomyelitis in one study,[78] but preoperative MRI had poor sensitivity and near-perfect specificity in another study of patients with septic hip.[98]

Inter-rater reliability and implementation

Magnetic resonance imaging appeared to have acceptable inter-rater reliability. There was some evidence suggesting that PET and scintigraphy had near-perfect inter-rater reliability, although this is limited to few small studies.[58,59,92] We found no evidence on patient preferences and cost-effectiveness of imaging tests for osteomyelitis.

Strengths and limitations of the assessment

Strengths

This systematic review was conducted in accordance with PRISMA guidelines; searches of numerous databases were performed, studies were assessed for quality using the QUADAS-2 framework and, provided they reported sufficient data, studies were synthesised using bivariate meta-analysis methods to assess diagnostic accuracy.

Over 18,000 database records were assessed and 81 published studies were included in this review, making it the largest and most comprehensive review of the diagnosis of osteomyelitis to date. This was the first review in the field to consider all relevant imaging tests across all types of patient receiving imaging tests to diagnose osteomyelitis. This was, therefore, the first review to our knowledge that was able to comprehensively compare imaging tests and to investigate whether or not the choice of imaging test varies with type of patient.

This review used recommended contemporary meta-analysis methods such as bivariate analysis and HSROC analysis to synthesise studies, so the synthesis is robust and meets the highest statistical standards. This was the first meta-analysis of osteomyelitis and, to our knowledge, the first practical diagnostic meta-analysis in any field, to use regression models to formally compare different diagnostic tests. This made it the first analysis to comprehensively and reliably compare the various imaging tests for osteomyelitis.

Limitations

The limitations of this review are largely a consequence of the limitations in the identified studies.

There were numerous concerns about the potential for bias in the included studies. Most studies were small, with < 50 participants, and were conducted retrospectively. The risk-of-bias assessment suggested potential bias because of unclear methods of patient selection and lack of blinding between index tests and references standards. However, statistical analysis found no evidence that these concerns led to actual biases in the results.

There were few studies identified that included children. This may be because studies of children do not discuss osteomyelitis directly, but only in the context of other conditions such as septic arthritis. Hence, it is possible that some possibly relevant studies were missed. However, subsequent informal searches did not identify any further useful reviews or studies that could be included and so we were unable to draw any firm conclusions on the accuracy of imaging in children.

Some imaging tests were reported in a few studies, particularly ultrasound and CT, so we were not able to fully assess their diagnostic accuracy. Some aspects of studies were inconsistently reported, such as varying descriptions of the cause of osteomyelitis, non-reporting of whether osteomyelitis was acute or chronic, and lack of clarity on whether or not radiography (or other tests) had been used prior to the main test. This made assessment in these subgroups difficult.

We identified very few data beyond those on diagnostic accuracy, with few studies discussing broader implementation issues such as access to machinery, costs or radiation exposure. This meant that this review could not consider the broader issues of how imaging tests should be used, for example whether or not differences in diagnostic accuracy between PET/SPECT and MRI merit the additional radiation exposure, or suggest that PET/SPECT machines should be made more widely available.

Scintigraphy was not specified in the protocol as a test of interest as it is little used in the UK. However, many studies of scintigraphy were identified during abstract screening and many studies compare scintigraphy with other tests. Given this, we decided to include scintigraphy in the main analysis where it had been compared with other tests. Studies of scintigraphy alone were included in further analyses, but these studies were not fully assessed for quality. This uncertainty in quality, and the consequence that the scintigraphy-only studies may be of lower quality than the studies that were fully assessed, is a limitation of this analysis.

Our analysis therefore has the potential for bias, either because the diagnostic accuracy of scintigraphy may be underestimated when it is compared with other tests expected to be better, or because studies

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

of scintigraphy alone may be of poorer quality. However, the diagnostic accuracy of the scintigraphy-only studies was similar to that of the fully assessed studies. Any bias is likely to be modest and the overall conclusions reasonable. The benefits of assessing all the evidence on scintigraphy (e.g. being able to compare different test types) probably outweigh any risk of bias.

Uncertainties

The main uncertainties remaining from this review arise largely because of limitations in the identified studies.

The diagnostic accuracy of imaging tests in children remains highly uncertain owing to the very limited nature of the evidence. We could reach no firm conclusions on the diagnostic accuracy of any imaging test in children.

The diagnostic accuracy of ultrasound is currently unknown as the only two studies in adults had conflicting results. However, DORs in both studies were low, suggesting that ultrasound is unlikely to have similar or better diagnostic accuracy than MRI, PET or SPECT. Ultrasound, however, is a test of choice in patients with metalwork in situ as this causes artefacts in CT and MRI.

Although we found no evidence that diagnostic accuracy varied across subgroups of patients, limited or inconsistent reporting of some characteristics, such as AOM versus COM, or the cause or location of osteomyelitis, means than differences between tests or between subgroups cannot be ruled out.

Considerations from patient representatives

Both patient representatives were parents of children who were treated for osteomyelitis. Both discussed the fact that diagnosis was generally a long process, with multiple imaging tests performed, with preliminary radiography and then ultrasound tests, both of which might be repeated. MRI was used for the 'definitive' diagnosis, and MRI was also used to confirm that treatment had been effective. We note that this process of multiple tests and their interactions could not be easily captured in this systematic review, and the impact and value of multiple imaging tests remain uncertain.

Antibiotics were given early in the diagnostic process, before ultrasound or MRI was performed. This may affect evaluations of diagnostic accuracy, because osteomyelitis is treated before being formally diagnosed. In future studies, the impact on diagnostic accuracy of treatment, given the ethical need for prompt treatment, will need to be considered.

Both parents noted that there was little time to discuss the tests or decide which was best for their child and they accepted whatever their doctors recommended. Both felt that the need to correctly diagnose osteomyelitis outweighed any concerns about the tests themselves. So, for example, they were willing to accept MRI requiring general anaesthesia.

Chapter 6 Conclusions

Implications for health care

This review found that MRI, PET and SPECT all have broadly similar and high accuracy when diagnosing osteomyelitis. All three tests correctly diagnose most people with, and most people without, osteomyelitis. All three are therefore suitable imaging tests for reliably diagnosing osteomyelitis. No clear reason to prefer one test over the other in terms of diagnostic accuracy was identified. The wider availability of MRI machines, and the fact that MRI does not expose patients to harmful ionising radiation, may mean that MRI is preferable in most cases, unless it is unsuitable for a particular patient. MRI also gives additional information as to the location of fluid collections and anatomical detail for surgical planning not given by SPECT or PET in such detail. PET or SPECT may be required if MRI is inconclusive.

Magnetic resonance imaging generally had poorer specificity than PET or SPECT, with a wide range of specificities across studies. This means that there may be substantial numbers of 'false-positive' cases when using MRI and clinicians should be aware of this potential for overdiagnosis. This may mean that MRI may be best suited to cases where false-positive cases are of lesser concern, perhaps such as when patients would proceed directly to antibiotic treatment. This might, however, be an artefact of the diagnostic studies; actual diagnoses of osteomyelitis will consider the clinical background of the patient (e.g. fever and infection markers in the blood test), so diagnosis is not based on the imaging test alone.

Position emission tomography had poorer sensitivity, but higher specificity, than MRI, with more consistent results across studies. This may make PET better suited to situations where avoiding false-positive diagnoses is important, for example when the test would be followed by surgery or other invasive procedures. However, MRI may still provide greater visual detail to guide the surgical approach than PET/CT alone.

Scintigraphy, in general, is less accurate than MRI or PET, but current methods of scintigraphy, particularly ^{99m}Tc HMPAO WBC scintigraphy, may have comparable accuracy, and this should be the preferred approach. However, planar scintigraphy may have poorer diagnostic accuracy than 3D SPECT imaging, so should perhaps be used only when SPECT is unavailable or unsuitable.

Data on ultrasound and CT were limited, but suggested that these tests have poorer diagnostic accuracy and perhaps should not currently be used in adults.

Radiography on its own has poor diagnostic accuracy so should probably not be used in isolation. This review could not confirm whether or not using radiography prior to MRI or other tests improves accuracy or reduces costs. It is, therefore, unclear whether or not the common practice of using radiography as a first test and proceeding to a MRI or another test if the radiography is inconclusive maximises diagnostic accuracy. It is, however, likely that radiography can be usefully used to rule out osteomyelitis in more obvious cases (e.g. when symptoms are the result of bone fracture). Radiography is probably more useful in chronic or subacute osteomyelitis, when bone destruction is already visible, but not in acute haematogenous osteomyelitis, when early changes are looked for.

The review found that diagnostic accuracy in adults does not appear to vary with the potential cause of osteomyelitis, or with the body part scanned. In particular, there was no evidence that diagnostic accuracy was different in patients with diabetic foot ulcers. This suggests that MRI, PET and SPECT are all reasonable tests to use in adults, regardless of the underlying conditions the patient has. The choice may depend primarily on suitability for the patient, such as potential difficulties in interpreting MRI scans in patients with indwelling metalwork. However, data on specific locations or causes of osteomyelitis other than diabetic foot ulcers were limited, so the possibility that diagnostic accuracy is different for some locations or causes remains.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

The review identified very limited data on diagnosing osteomyelitis in children, so considerable uncertainty remains over the diagnostic accuracy of imaging tests in children. Clinicians should be aware of this limitation in the evidence base.

There was some evidence to suggest that MRI has an acceptable degree of agreement among radiologists. The evidence for the inter-rater reliability of other imaging tests is too limited to draw conclusions. There is also very limited evidence on the implementation of imaging tests for the diagnosis of osteomyelitis. Notably, there is a lack of evidence on patient preferences and on the most cost-effective diagnostic imaging pathways for osteomyelitis.

Suggested research priorities

The most urgent research priority is to perform diagnostic accuracy studies of imaging tests in children, given the limitations of the current evidence base. Large diagnostic accuracy studies are needed to assess imaging tests in children. These should be of high quality, with proper blinding of test assessors and consecutive recruitment of patients, and reporting of the appropriate reference standard used to confirm positive and negative diagnoses. The priority tests should be MRI and ultrasound, ideally comparing the two tests in the same children.

Ultrasound has not been widely assessed in adults. Current results suggest that ultrasound on its own may not be sufficiently accurate to diagnose osteomyelitis, but further accuracy studies are needed to resolve the uncertainty. It may be more appropriate to investigate the diagnostic accuracy of ultrasound as a precursor to MRI or other tests (e.g. as a replacement for radiography).

Consistency of test interpretation is an important area in diagnostic testing that is often overlooked. Further investigation of inter- and intra-rater reliability for all tests is needed. Ideally this should be included in diagnostic accuracy tests, but it could also be investigated in case series or clinical audits.

Given the similarities in diagnostic accuracy of MRI, PET and SPECT, suitable investigation of patient and clinician experience and opinion of these tests, through surveys or focus groups, would be useful to identify practical reasons for the choice of test. Similarly, a formal economic evaluation of these tests, accounting for test cost, sequencing of tests, availability and risk of radiation exposure, would also help to clarify the choice between these tests.

Acknowledgements

We would like to acknowledge Teumzghi Mebrahtu, Alex Hodgkinson (both formerly of CRD) and Gary Raine (CRD) for their work on abstract screening and data extraction for this review; Nerys Woolacott (formerly of CRD) for her support in establishing the project and creating the protocol; Andrew Grainger, Emily Hargreaves and Sean O'Riordan (Leeds Teaching Hospitals) for their clinical advice over the course of the project; and Rachel McDonald and Ryan Harrison for their support on providing patient and carer perspectives and experience.

Contributions of authors

Alexis Llewellyn (https://orcid.org/0000-0003-4569-5136) performed the systematic review, including screening, study selection, data extraction, quality assessment and narrative synthesis. He wrote most sections of this report.

Julie Jones-Diette (https://orcid.org/0000-0003-1769-8612) performed data extraction and quality assessment and prepared data tables for the report.

Jeannette Kraft (https://orcid.org/0000-0003-2705-1302) and **Colin Holton** provided clinical expertise and commented on drafts of the protocol and this report.

Melissa Harden prepared and performed all database searches for the systematic review.

Mark Simmonds (https://orcid.org/0000-0002-1999-8515) oversaw the project as a whole, performed the statistical analyses and wrote all sections of the report pertaining to those analyses.

Data-sharing statement

This is a systematic review of published data, and no primary data were created for this report. All data used are as reported in the publications listed throughout this report. Data-extraction forms are available on request from the corresponding author.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

References

1. National Institute for Health and Care Excellence (NICE). *Diabetic Foot Problems: Prevention and Management*. NICE Guideline [NG19]. London: NICE; 2015. URL: www.nice.org.uk/guidance/ng19 (accessed 6 August 2019).
2. Malhotra R, Chan CS, Nather A. Osteomyelitis in the diabetic foot. *Diabet Foot Ankle* 2014;**5**:1. https://doi.org/10.3402/dfa.v5.24445
3. Josse J, Velard F, Gangloff SC. *Staphylococcus aureus* vs. osteoblast: relationship and consequences in osteomyelitis. *Front Cell Infect Microbiol* 2015;**5**:85. https://doi.org/10.3389/fcimb.2015.00085
4. Kremers HM, Nwojo ME, Ransom JE, Wood-Wentz CM, Melton LJ, Huddleston PM. Trends in the epidemiology of osteomyelitis: a population-based study, 1969 to 2009. *J Bone Joint Surg Am* 2015;**97**:837–45. https://doi.org/10.2106/JBJS.N.01350
5. Lew DP, Waldvogel FA. Osteomyelitis. *N Engl J Med* 1997;**336**:999–1007. https://doi.org/10.1056/NEJM199704033361406
6. Lew DP, Waldvogel FA. Osteomyelitis. *Lancet* 2004;**364**:369–79. https://doi.org/10.1016/S0140-6736(04)16727-5
7. Berendt AR, Peters EJ, Bakker K, Embil JM, Eneroth M, Hinchliffe RJ, *et al.* Diabetic foot osteomyelitis: a progress report on diagnosis and a systematic review of treatment. *Diabetes Metab Res Rev* 2008;**24**(Suppl. 1):145–61. https://doi.org/10.1002/dmrr.836
8. Lima AL, Oliveira PR, Carvalho VC, Cimerman S, Savio E, Diretrizes Panamericanas para el Tratamiento de las Osteomielitis e Infecciones de Tejidos Blandos Group. Recommendations for the treatment of osteomyelitis. *Braz J Infect Dis* 2014;**18**:526–34. https://doi.org/10.1016/j.bjid.2013.12.005
9. Carek PJ, Dickerson LM, Sack JL. Diagnosis and management of osteomyelitis. *Am Fam Physician* 2001;**63**:2413–20.
10. Gold R. Diagnosis of osteomyelitis. *Pediatr Rev* 1991;**12**:292–7. https://doi.org/10.1542/pir.12-10-292
11. Karmazyn B. Imaging approach to acute hematogenous osteomyelitis in children: an update. *Semin Ultrasound CT MR* 2010;**31**:100–6. https://doi.org/10.1053/j.sult.2009.12.002
12. Pineda C, Espinosa R, Pena A. Radiographic imaging in osteomyelitis: the role of plain radiography, computed tomography, ultrasonography, magnetic resonance imaging, and scintigraphy. *Semin Plast Surg* 2009;**23**:80–9. https://doi.org/10.1055/s-0029-1214160
13. Papanas N, Zissimopoulos A, Maltezos E. (18)F-FDG PET and PET/CT for the diagnosis of diabetic foot osteomyelitis. *Hippokratia* 2013;**17**:4–6.
14. van Schuppen J, van Doorn MM, van Rijn RR. Childhood osteomyelitis: imaging characteristics. *Insights Imaging* 2012;**3**:519–33. https://doi.org/10.1007/s13244-012-0186-8
15. Centre for Clinical Practice at National Institute for Health and Clinical Excellence (NICE). Diabetic Foot Problems: Inpatient Management of Diabetic Foot Problems. London: NICE; 2011.
16. Hatzenbuehler J, Pulling TJ. Diagnosis and management of osteomyelitis. *Am Fam Physician* 2011;**84**:1027–33.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

17. Lipsky BA, Berendt AR, Cornia PB, Pile JC, Peters EJ, Armstrong DG, *et al.* 2012 Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. *Clin Infect Dis* 2012;**54**:e132–73. https://doi.org/10.1093/cid/cis346

18. Conterno LO, da Silva Filho CR. Antibiotics for treating chronic osteomyelitis in adults. *Cochrane Database Syst Rev* 2009;**3**:CD004439. https://doi.org/10.1002/14651858.CD004439.pub2

19. Stengel D, Bauwens K, Sehouli J, Ekkernkamp A, Porzsolt F. Systematic review and meta-analysis of antibiotic therapy for bone and joint infections. *Lancet Infect Dis* 2001;**1**:175–88. https://doi.org/10.1016/S1473-3099(01)00094-9

20. Dinh MT, Abad CL, Safdar N. Diagnostic accuracy of the physical examination and imaging tests for osteomyelitis underlying diabetic foot ulcers: meta-analysis. *Clin Infect Dis* 2008;**47**:519–27. https://doi.org/10.1086/590011

21. Govaert GA, IJpma FF, McNally M, McNally E, Reininga IH, Glaudemans AW. Accuracy of diagnostic imaging modalities for peripheral post-traumatic osteomyelitis – a systematic review of the recent literature. *Eur J Nucl Med Mol Imaging* 2017;**44**:1393–407. https://doi.org/10.1007/s00259-017-3683-7

22. Kapoor A, Page S, Lavalley M, Gale DR, Felson DT. Magnetic resonance imaging for diagnosing foot osteomyelitis: a meta-analysis. *Arch Intern Med* 2007;**167**:125–32. https://doi.org/10.1001/archinte.167.2.125

23. Lauri C, Tamminga M, Glaudemans AWJM, Juárez Orozco LE, Erba PA, Jutte PC, *et al.* Detection of osteomyelitis in the diabetic foot by imaging techniques: a systematic review and meta-analysis comparing MRI, white blood cell scintigraphy, and FDG-PET. *Diabetes Care* 2017;**40**:1111–20. https://doi.org/10.2337/dc17-0532

24. Mooney ML, Haidet K, Liu J, Ebraheim NA. Hematogenous calcaneal osteomyelitis in children. *Foot Ankle Spec* 2017;**10**:63–8. https://doi.org/10.1177/1938640016679704

25. Termaat MF, Raijmakers PG, Scholten HJ, Bakker FC, Patka P, Haarman HJ. The accuracy of diagnostic imaging for the assessment of chronic osteomyelitis: a systematic review and meta-analysis. *J Bone Joint Surg Am* 2005;**87**:2464–71. https://doi.org/10.2106/00004623-200511000-00013

26. Treglia G, Sadeghi R, Annunziata S, Zakavi SR, Caldarella C, Muoio B, *et al.* Diagnostic performance of fluorine-18-fluorodeoxyglucose positron emission tomography for the diagnosis of osteomyelitis related to diabetic foot: a systematic review and a meta-analysis. *Foot* 2013;**23**:140–8. https://doi.org/10.1016/j.foot.2013.07.002

27. van der Bruggen W, Bleeker-Rovers CP, Boerman OC, Gotthardt M, Oyen WJ. PET and SPECT in osteomyelitis and prosthetic bone and joint infections: a systematic review. *Semin Nucl Med* 2010;**40**:3–15. https://doi.org/10.1053/j.semnuclmed.2009.08.005

28. Wang GL, Zhao K, Liu ZF, Dong MJ, Yang SY. A meta-analysis of fluorodeoxyglucose-positron emission tomography versus scintigraphy in the evaluation of suspected osteomyelitis. *Nucl Med Commun* 2011;**32**:1134–42. https://doi.org/10.1097/MNM.0b013e32834b455c

29. Xu Z, Koo A, Shah A. The utility of MRI for the diagnosis of osteomyelitis in the pressure ulcer patient. *J Plast Reconstr Aesthet Surg* 2017;**70**:289–91. https://doi.org/10.1016/j.bjps.2016.11.009

30. Dartnell J, Ramachandran M, Katchburian M. Haematogenous acute and subacute paediatric osteomyelitis: a systematic review of the literature. *J Bone Joint Surg Br* 2012;**94**:584–95. https://doi.org/10.1302/0301-620X.94B5.28523

31. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

32. van de Pol RJ, van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. *J Physiother* 2010;**56**:7–17. https://doi.org/10.1016/S1836-9553(10)70049-7

33. Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res* 2016;**25**:2858–77. https://doi.org/10.1177/0962280214534409

34. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90. https://doi.org/10.1016/j.jclinepi.2005.02.022

35. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;**20**:2865–84. https://doi.org/10.1002/sim.942

36. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60. https://doi.org/10.1136/bmj.327.7414.557

37. Abdel Razek AAK, Samir S. Diagnostic performance of diffusion-weighted MR imaging in differentiation of diabetic osteoarthropathy and osteomyelitis in diabetic foot. *Eur J Radiol* 2017;**89**:221–5. https://doi.org/10.1016/j.ejrad.2017.02.015

38. Al-Khawari HA, Al-Saeed OM, Jumaa TH, Chishti F. Evaluating diabetic foot infection with magnetic resonance imaging: Kuwait experience. *Med Princ Pract* 2005;**14**:165–72. https://doi.org/10.1159/000084634

39. Al-Sheikh W, Sfakianakis GN, Mnaymneh W, Hourani M, Heal A, Duncan RC, *et al.* Subacute and chronic bone infections: diagnosis using In-111, ^{67}Ga and Tc-99m MDP bone scintigraphy, and radiography. *Radiology* 1985;**155**:501–6. https://doi.org/10.1148/radiology.155.2.3157204

40. Aragón-Sánchez J, Lipsky BA, Lazaro-Martinez JL. Diagnosing diabetic foot osteomyelitis: is the combination of probe-to-bone test and plain radiography sufficient for high-risk inpatients? *Diabet Med* 2011;**28**:191–4. https://doi.org/10.1111/j.1464-5491.2010.03150.x

41. Aslangul E, M'bemba J, Caillat-Vigneron N, Coignard S, Larger E, Boitard C, Lipsky BA. Diagnosing diabetic foot osteomyelitis in patients without signs of soft tissue infection by coupling hybrid 67Ga SPECT/CT with bedside percutaneous bone puncture. *Diabetes Care* 2013;**36**:2203–10. https://doi.org/10.2337/dc12-2108

42. Blume PA, Dey HM, Daley LJ, Arrighi JA, Soufer R, Gorecki GA. Diagnosis of pedal osteomyelitis with Tc-99m HMPAO labeled leukocytes. *J Foot Ankle Surg* 1997;**36**:120–6. https://doi.org/10.1016/S1067-2516(97)80057-9

43. Bohchelian HA, Klisarova AD, Koeva LA. Single photon emission tomography in radioimmune imaging of diabetic foot infection. *Diabetol Pol* 2002;**9**:39–44.

44. Bolouri C, Merwald M, Huellner MW, Veit-Haibach P, Kuttenberger J, Pérez-Lago M, *et al.* Performance of orthopantomography, planar scintigraphy, CT alone and SPECT/CT in patients with suspected osteomyelitis of the jaw. *Eur J Nucl Med Mol Imaging* 2013;**40**:411–17. https://doi.org/10.1007/s00259-012-2285-7

45. Brunel AS, Lamy B, Cyteval C, Perrochia H, Téot L, Masson R, *et al.* Diagnosing pelvic osteomyelitis beneath pressure ulcers in spinal cord injured patients: a prospective study. *Clin Microbiol Infect* 2016;**22**:267.e1–8. https://doi.org/10.1016/j.cmi.2015.11.005

46. Chacko TK, Zhuang H, Nakhoda KZ, Moussavian B, Alavi A. Applications of fluorodeoxyglucose positron emission tomography in the diagnosis of infection. *Nucl Med Commun* 2003;**24**:615–24. https://doi.org/10.1097/01.mnm.0000075189.60210.df

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

47. Croll SD, Nicholas GG, Osborne MA, Wasser TE, Jones S. Role of magnetic resonance imaging in the diagnosis of osteomyelitis in diabetic foot infections. *J Vasc Surg* 1996;**24**:266–70. https://doi.org/10.1016/S0741-5214(96)70102-7

48. Demirev A, Weijers R, Geurts J, Mottaghy F, Walenkamp G, Brans B. Comparison of [18 F]FDG PET/CT and MRI in the diagnosis of active osteomyelitis. *Skeletal Radiol* 2014;**43**:665–72. https://doi.org/10.1007/s00256-014-1844-3

49. Enderle MD, Coerper S, Schweizer HP, Kopp AE, Thelen MH, Meisner C, *et al.* Correlation of imaging techniques to histopathology in patients with diabetic foot syndrome and clinical suspicion of chronic osteomyelitis. The role of high-resolution ultrasound. *Diabetes Care* 1999;**22**:294–9. https://doi.org/10.2337/diacare.22.2.294

50. Erdman WA, Tamburro F, Jayson HT, Weatherall PT, Ferry KB, Peshock RM. Osteomyelitis: characteristics and pitfalls of diagnosis with MR imaging. *Radiology* 1991;**180**:533–9. https://doi.org/10.1148/radiology.180.2.2068324

51. Ertugrul MB, Baktiroglu S, Salman S, Unal S, Aksoy M, Berberoglu K, Calangu S. The diagnosis of osteomyelitis of the foot in diabetes: microbiological examination vs. magnetic resonance imaging and labelled leucocyte scanning. *Diabet Med* 2006;**23**:649–53. https://doi.org/10.1111/j.1464-5491.2006.01887.x

52. Ezzat T, El-Hamid AA, Mostafa M, El-Kady L. Early diagnosis of acute osteomyelitis in children by high-resolution and power Doppler sonography. *Egypt J Radiol Nucl Med* 2011;**42**:233–42. https://doi.org/10.1016/j.ejrnm.2011.05.002

53. Familiari D, Glaudemans AW, Vitale V, Prosperi D, Bagni O, Lenza A, *et al.* Can sequential 18F-FDG PET/CT replace WBC imaging in the diabetic foot? *J Nucl Med* 2011;**52**:1012–19. https://doi.org/10.2967/jnumed.110.082222

54. Filippi L, Schillaci O. Usefulness of hybrid SPECT/CT in 99mTc-HMPAO-labeled leukocyte scintigraphy for bone and joint infections. *J Nucl Med* 2006;**47**:1908–13.

55. Filippi L, Uccioli L, Giurato L, Schillaci O. Diabetic foot infection: usefulness of SPECT/CT for 99mTc-HMPAO-labeled leukocyte imaging. *J Nucl Med* 2009;**50**:1042–6. https://doi.org/10.2967/jnumed.108.059493

56. Franceschi D, Matthews R, Brunetti V, Franceschi AM, Safaie E, Carcamo CJ, *et al.* FDG PET/CT imaging in diabetic foot infections. *Eur J Nucl Med Mol Imaging* 2013;**40**(Suppl. 2):1–477. https://doi.org/10.1007/s00259-013-2535-3

57. Gemmel F, De Winter F, Van Laere K, Vogelaers D, Uyttendaele D, Dierckx RA. 99mTc ciprofloxacin imaging for the diagnosis of infection in the postoperative spine. *Nucl Med Commun* 2004;**25**:277–83. https://doi.org/10.1097/00006231-200403000-00011

58. Guhlmann A, Brecht-Krauss D, Suger G, Glatting G, Kotzerke J, Kinzl L, Reske SN. Fluorine-18-FDG PET and technetium-99m antigranulocyte antibody scintigraphy in chronic osteomyelitis. *J Nucl Med* 1998;**39**:2145–52.

59. Guhlmann A, Brecht-Krauss D, Suger G, Glatting G, Kotzerke J, Kinzl L, Reske SN. Chronic osteomyelitis: detection with FDG PET and correlation with histopathologic findings. *Radiology* 1998;**206**:749–54. https://doi.org/10.1148/radiology.206.3.9494496

60. Hakim SG, Bruecker CW, Jacobsen HCh, Hermes D, Lauer I, Eckerle S, *et al.* The value of FDG-PET and bone scintigraphy with SPECT in the primary diagnosis and follow-up of patients with chronic osteomyelitis of the mandible. *Int J Oral Maxillofac Surg* 2006;**35**:809–16. https://doi.org/10.1016/j.ijom.2006.03.029

61. Hartmann A, Eid K, Dora C, Trentz O, von Schulthess GK, Stumpe KDM. Diagnostic value of ^{18}F-FDG PET/CT in trauma patients with suspected chronic osteomyelitis. *Eur J Nucl Med Mol Imaging* 2007;**34**:704–14. https://doi.org/10.1007/s00259-006-0290-4

62. Hazenberg C, Kraai M, Moll F, van Baal S. Prediction of outcome of treatment of osteomyelitis in diabetic foot after minor amputation: the role of MR imaging. *J Vasc Surg* 2011;**53**:78S–79S. http://dx.doi.org/10.1016/j.jvs.2011.03.151

63. Heiba SI, Stempler L, Sullivan T, Kolker D, Kostakoglu L. The ideal dual-isotope imaging combination in evaluating patients with suspected infection of pelvic pressure ulcers. *Nucl Med Commun* 2017;**38**:129–34. https://doi.org/10.1097/MNM.0000000000000625

64. Horger M, Eschmann SM, Pfannenberg C, Storek D, Dammann F, Vonthein R, *et al.* The value of SPET/CT in chronic osteomyelitis. *Eur J Nucl Med Mol Imaging* 2003;**30**:1665–73. https://doi.org/10.1007/s00259-003-1321-z

65. Horger M, Eschmann SM, Pfannenberg C, Storek D, Vonthein R, Claussen CD, Bares R. Added value of SPECT/CT in patients suspected of having bone infection: preliminary results. *Arch Orthop Trauma Surg* 2007;**127**:211–21. https://doi.org/10.1007/s00402-006-0259-6

66. Huang AB, Schweitzer ME, Hume E, Batte WG. Osteomyelitis of the pelvis/hips in paralyzed patients: accuracy and clinical utility of MRI. *J Comput Assist Tomogr* 1998;**22**:437–43. https://doi.org/10.1097/00004728-199805000-00017

67. Johnson PW, Collins MS, Wenger DE. Diagnostic utility of T1-weighted MRI characteristics in evaluation of osteomyelitis of the foot. *AJR Am J Roentgenol* 2009;**192**:96–100. https://doi.org/10.2214/AJR.08.1376

68. Kaim A, Ledermann HP, Bongartz G, Messmer P, Müller-Brand J, Steinbrich W. Chronic post-traumatic osteomyelitis of the lower extremity: comparison of magnetic resonance imaging and combined bone scintigraphy/immunoscintigraphy with radiolabelled monoclonal antigranulocyte antibodies. *Skeletal Radiol* 2000;**29**:378–86. https://doi.org/10.1007/s002560000228

69. Kim C, Lee SJ, Kim JY, Hwang KT, Choi YY. Comparative analysis of 99mTc-MDP three-phase bone scan with SPECT/CT and 99mTc-HMPAO-Labeled WBC SPECT/CT in the differential diagnosis of clinically suspicious post-traumatic osteomyelitis. *Nucl Med Mol Imaging* 2017;**51**:40–8. https://doi.org/10.1007/s13139-016-0441-x

70. La Fontaine J, Bhavan K, Lam K, Van Asten S, Erdman W, Lavery LA, Öz OK. Comparison between Tc-99m WBC SPECT/CT and MRI for the diagnosis of biopsy-proven diabetic foot osteomyelitis. *Wounds* 2016;**28**:271–8.

71. Larcos G, Brown ML, Sutton RT. Diagnosis of osteomyelitis of the foot in diabetic patients: value of ^{111}In-leukocyte scintigraphy. *AJR Am J Roentgenol* 1991;**157**:527–31. https://doi.org/10.2214/ajr.157.3.1872240

72. Larson DL, Gilstrap J, Simonelic K, Carrera GF. Is there a simple, definitive, and cost-effective way to diagnose osteomyelitis in the pressure ulcer patient? *Plast Reconstr Surg* 2011;**127**:670–6. https://doi.org/10.1097/PRS.0b013e3181fed66e

73. Ledermann HP, Kaim A, Bongartz G, Steinbrich W. Pitfalls and limitations of magnetic resonance imaging in chronic posttraumatic osteomyelitis. *Eur Radiol* 2000;**10**:1815–23. https://doi.org/10.1007/s003300000480

74. Levine SE, Neagle CE, Esterhai JL, Wright DG, Dalinka MK. Magnetic resonance imaging for the diagnosis of osteomyelitis in the diabetic patient with a foot ulcer. *Foot Ankle Int* 1994;**15**:151–6. https://doi.org/10.1177/107110079401500311

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

75. Lewis VL, Bailey MH, Pulawski G, Kind G, Bashioum RW, Hendrix RW. The diagnosis of osteomyelitis in patients with pressure sores. *Plast Reconstr Surg* 1988;**81**:229–32. https://doi.org/10.1097/00006534-198802000-00016

76. Lipman BT, Collier BD, Carrera GF, Timins ME, Erickson SJ, Johnson JE, *et al.* Detection of osteomyelitis in the neuropathic foot: nuclear medicine, MRI and conventional radiography. *Clin Nucl Med* 1998;**23**:77–82. https://doi.org/10.1097/00003072-199802000-00003

77. Mahendra M, Singh R. Diagnostic accuracy and surgical utility of MRI in complicated diabetic foot. *J Clin Diagn Res* 2017;**11**:RC01–RC04. https://doi.org/10.7860/JCDR/2017/25902.10154

78. Malcius D, Jonkus M, Kuprionis G, Maleckas A, Monastyreckiene E, Uktveris R, *et al.* The accuracy of different imaging techniques in diagnosis of acute hematogenous osteomyelitis. *Medicina* 2009;**45**:624–31. https://doi.org/10.3390/medicina45080081

79. Mason MD, Zlatkin MB, Esterhai JL, Dalinka MK, Velchik MG, Kressel HY. Chronic complicated osteomyelitis of the lower extremity: evaluation with MR imaging. *Radiology* 1989;**173**:355–9. https://doi.org/10.1148/radiology.173.2.2798868

80. McCarthy J, Hartmann E, Bentz ML, Rao VK, Jee Y, Rivedal D, Poore SO. Seeing is believing? Preoperative magnetic resonance imaging for pressure ulcers: implications for surgical management. *Plast Reconstr Surg Glob Open* 2017;**5**:e1263. https://doi.org/10.1097/GOX.0000000000001263

81. Meller J, Koster G, Liersch T, Siefker U, Lehmann K, Meyer I, *et al.* Chronic bacterial osteomyelitis: prospective comparison of ^{18}F-FDG imaging with a dual-head coincidence camera and ^{111}In-labelled autologous leucocyte scintigraphy. *Eur J Nucl Med Mol Imaging* 2002;**29**:53–60. https://doi.org/10.1007/s00259-001-0661-9

82. Miki F, Armstrong DG, Hiroto T. Efficacy of magnetic resonance imaging in deciding the appropriate surgical margin in diabetic foot osteomyelitis. *EWMA J* 2015;**15**:8–12.

83. Modic MT, Feiglin DH, Piraino DW, Boumphrey F, Weinstein MA, Duchesneau PM, Rehm S. Vertebral osteomyelitis: assessment using MR. *Radiology* 1985;**157**:157–66. https://doi.org/10.1148/radiology.157.1.3875878

84. Morales Lozano R, González Fernández ML, Martinez Hernández D, Beneit Montesinos JV, Guisado Jiménez S, Gonzalez Jurado MA. Validating the probe-to-bone test and other tests for diagnosing chronic osteomyelitis in the diabetic foot. *Diabetes Care* 2010;**33**:2140–5. https://doi.org/10.2337/dc09-2309

85. Morrison WB, Schweitzer ME, Bock GW, Mitchell DG, Hume EL, Pathria MN, Resnick D. Diagnosis of osteomyelitis: utility of fat-suppressed contrast-enhanced MR imaging. *Radiology* 1993;**189**:251–7. https://doi.org/10.1148/radiology.189.1.8204132

86. Morrison WB, Schweitzer ME, Batte WG, Radack DP, Russel KM. Osteomyelitis of the foot: relative importance of primary and secondary MR imaging signs. *Radiology* 1998;**207**:625–32. https://doi.org/10.1148/radiology.207.3.9609883

87. Nath AK, Sethu AU. Use of ultrasound in osteomyelitis. *Br J Radiol* 1992;**65**:649–52. https://doi.org/10.1259/0007-1285-65-776-649

88. Nawaz A, Torigian DA, Siegelman ES, Basu S, Chryssikos T, Alavi A. Diagnostic performance of FDG-PET, MRI, and plain film radiography (PFR) for the diagnosis of osteomyelitis in the diabetic foot. *Mol Imaging Biol* 2010;**12**:335–42. https://doi.org/10.1007/s11307-009-0268-2

89. Newman LG, Waller J, Palestro CJ, Schwartz M, Klein MJ, Hermann G, *et al.* Unsuspected osteomyelitis in diabetic foot ulcers. Diagnosis and monitoring by leukocyte scanning with indium in 111 oxyquinoline. *JAMA* 1991;**266**:1246–51. https://doi.org/10.1001/jama.1991.03470090080036

90. Newman LG, Waller J, Palestro CJ, Hermann G, Klein MJ, Schwartz M, *et al.* Leukocyte scanning with ^{111}In is superior to magnetic resonance imaging in diagnosis of clinically unsuspected osteomyelitis in diabetic foot ulcers. *Diabetes Care* 1992;**15**:1527–30. https://doi.org/10.2337/diacare.15.11.1527

91. Nigro ND, Bartynski WS, Grossman SJ, Kruljac S. Clinical impact of magnetic resonance imaging in foot osteomyelitis. *J Am Podiatr Med Assoc* 1992;**82**:603–15. https://doi.org/10.7547/87507315-82-12-603

92. Park HM, Wheat LJ, Siddiqui AR, Burt RW, Robb JA, Ransburg RC, Kernek CB. Scintigraphic evaluation of diabetic osteomyelitis: concise communication. *J Nucl Med* 1982;**23**:569–73.

93. Rastogi A, Bhattacharya A, Prakash M, Sharma S, Mittal BR, Khandelwal N, Bhansali A. Utility of PET/CT with fluorine-18-fluorodeoxyglucose-labeled autologous leukocytes for diagnosing diabetic foot osteomyelitis in patients with Charcot's neuroarthropathy. *Nucl Med Commun* 2016;**37**:1253–9. https://doi.org/10.1097/MNM.0000000000000603

94. Remedios D, Valabhji J, Oelbaum R, Sharp P, Mitchell R. 99mTc-nanocolloid scintigraphy for assessing osteomyelitis in diabetic neuropathic feet. *Clin Radiol* 1998;**53**:120–5. https://doi.org/10.1016/S0009-9260(98)80058-5

95. Rozzanigo U, Tagliani A, Vittorini E, Pacchioni R, Brivio LR, Caudana R. Role of magnetic resonance imaging in the evaluation of diabetic foot with suspected osteomyelitis. *Radiol Med* 2009;**114**:121–32. https://doi.org/10.1007/s11547-008-0337-7

96. Şanlı Y, Ozkan ZG, Unal SN, Türkmen C, Kılıçoğlu O. The additional value of Tc 99m HMPAO white blood cell SPECT in the evaluation of bone and soft tissue infections. *Mol Imaging Radionucl Ther* 2011;**20**:7–13. https://doi.org/10.4274/MIRT.20.02

97. Sarikaya A, Aygit AC, Pekindil G. Utility of 99mTc dextran scintigraphy in diabetic patients with suspected osteomyelitis of the foot. *Ann Nucl Med* 2003;**17**:669–76. https://doi.org/10.1007/BF02984973

98. Schlung JE, Bastrom TP, Roocroft JH, Newton PO, Mubarak SJ, Upasani VV. Femoral neck aspiration aids in the diagnosis of osteomyelitis in children with septic hip. *J Pediatr Orthop* 2016;**38**:532–536. https://doi.org/10.1097/BPO.0000000000000868

99. Schwegler B, Stumpe KD, Weishaupt D, Strobel K, Spinas GA, von Schulthess GK, *et al.* Unsuspected osteomyelitis is frequent in persistent diabetic foot ulcer and better diagnosed by MRI than by ^{18}F-FDG PET or 99mTc-MOAB. *J Intern Med* 2008;**263**:99–106. https://doi.org/10.1111/j.1365-2796.2007.01877.x

100. Seabold JE, Flickinger FW, Kao SC, Gleason TJ, Kahn D, Nepola JV, Marsh JL. Indium-111-leukocyte/technetium-99m-MDP bone and magnetic resonance imaging: difficulty of diagnosing osteomyelitis in patients with neuropathic osteoarthropathy. *J Nucl Med* 1990;**31**:549–56.

101. Seabold JE, Simonson TM, Weber PC, Thompson BH, Harris KG, Rezai K, *et al.* Cranial osteomyelitis: diagnosis and follow-up with In-111 white blood cell and Tc-99m methylene diphosphonate bone SPECT, CT, and MR imaging. *Radiology* 1995;**196**:779–88. https://doi.org/10.1148/radiology.196.3.7644643

102. Segall GM, Nino-Murcia M, Jacobs T, Chang K. The role of bone scan and radiography in the diagnostic evaluation of suspected pedal osteomyelitis. *Clin Nucl Med* 1989;**14**:255–60. https://doi.org/10.1097/00003072-198904000-00003

103. Shemesh S, Kosashvili Y, Groshar D, Bernstine H, Sidon E, Cohen N, *et al.* The value of 18-FDG PET/CT in the diagnosis and management of implant-related infections of the tibia: a case series. *Injury* 2015;**46**:1377–82. https://doi.org/10.1016/j.injury.2015.03.002

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

104. Unger E, Moldofsky P, Gatenby R, Hartz W, Broder G. Diagnosis of osteomyelitis by MR imaging. *AJR Am J Roentgenol* 1988;**150**:605–10. https://doi.org/10.2214/ajr.150.3.605

105. van Vliet KE, de Jong VM, Termaat MF, Schepers T, van Eck-Smit BLF, Goslings JC, Schep NWL. FDG-PET/CT for differentiating between aseptic and septic delayed union in the lower extremity. *Arch Orthop Trauma Surg* 2018;**138**:189–94. https://doi.org/10.1007/s00402-017-2806-8

106. Weber PC, Seabold JE, Graham SM, Hoffmann HH, Simonson TM, Thompson BH. Evaluation of temporal and facial osteomyelitis by simultaneous In-WBC/Tc-99m-MDP bone SPECT scintigraphy and computed tomography scan. *Otolaryngol Head Neck Surg* 1995;**113**:36–41. https://doi.org/10.1016/S0194-5998(95)70142-7

107. Weinstein D, Wang A, Chambers R, Stewart CA, Motz HA. Evaluation of magnetic resonance imaging in the diagnosis of osteomyelitis in diabetic foot infections. *Foot Ankle* 1993;**14**:18–22. https://doi.org/10.1177/107110079301400104

108. Wenter V, Müller JP, Albert NL, Lehner S, Fendler WP, Bartenstein P, *et al.* The diagnostic value of [^{18}F]FDG PET for the detection of chronic osteomyelitis and implant-associated infection. *Eur J Nucl Med Mol Imaging* 2016;**43**:749–61. https://doi.org/10.1007/s00259-015-3221-4

109. Weon YC, Yang SO, Choi YY, Shin JW, Ryu JS, Shin MJ, *et al.* Use of Tc-99m HMPAO leukocyte scans to evaluate bone infection: incremental value of additional SPECT images. *Clin Nucl Med* 2000;**25**:519–26. https://doi.org/10.1097/00003072-200007000-00006

110. Williamson BR, Teates CD, Phillips CD, Croft BY. Computed tomography as a diagnostic aid in diabetic and other problem feet. *Clin Imaging* 1989;**13**:159–63. https://doi.org/10.1016/0899-7071(89)90101-0

111. Yang H, Zhuang H, Rubello D, Alavi A. Mild-to-moderate hyperglycemia will not decrease the sensitivity of ^{18}F-FDG PET imaging in the detection of pedal osteomyelitis in diabetic patients. *Nucl Med Commun* 2016;**37**:259–62. https://doi.org/10.1097/MNM.0000000000000434

112. Yuh WT, Corson JD, Baraniewski HM, Rezai K, Shamma AR, Kathol MH, *et al.* Osteomyelitis of the foot in diabetic patients: evaluation with plain film, 99mTc-MDP bone scintigraphy, and MR imaging. *AJR Am J Roentgenol* 1989;**152**:795–800. https://doi.org/10.2214/ajr.152.4.795

113. Zaiton F, Samir AM, Elkamash TH, Tawfik AM, Hadhoud KM. Evaluation of diabetic foot osteomyelitis using probe to bone test and magnetic resonance imaging and their impact on surgical intervention. *Egyptian Journal of Radiology and Nuclear Medicine* 2014;**45**:795–802. https://doi.org/10.1016/j.ejrnm.2014.04.015

114. Álvaro-Afonso FJ, Lázaro-Martínez JL, Aragón-Sánchez J, García-Morales E, Cecilia-Matilla A, Beneit-Montesinos JV. Interobserver and intraobserver reproducibility of plain X-rays in the diagnosis of diabetic foot osteomyelitis. *Int J Low Extrem Wounds* 2013;**12**:12–15. https://doi.org/10.1177/1534734612474304

115. Álvaro-Afonso FJ, Lázaro-Martínez JL, Aragón-Sánchez J, García-Morales E, García-Álvarez Y, Molines-Barroso RJ. Inter-observer reproducibility of diagnosis of diabetic foot osteomyelitis based on a combination of probe-to-bone test and simple radiography. *Diabetes Res Clin Pract* 2014;**105**:e3–5. https://doi.org/10.1016/j.diabres.2014.04.024

116. Averill LW, Hernandez A, Gonzalez L, Peña AH, Jaramillo D. Diagnosis of osteomyelitis in children: utility of fat-suppressed contrast-enhanced MRI. *AJR Am J Roentgenol* 2009;**192**:1232–8. https://doi.org/10.2214/AJR.07.3400

117. Hauptfleisch J, Meagher TM, Hughes RJ, Singh JP, Graham A, López de Heredia L. Interobserver agreement of magnetic resonance imaging signs of osteomyelitis in pelvic pressure ulcers in patients with spinal cord injury. *Arch Phys Med Rehabil* 2013;**94**:1107–11. https://doi.org/10.1016/j.apmr.2012.11.012

118. Govaert GA, Glaudemans AW, Ploegmakers JJ, Viddeleer AR, Wendt KW, Reininga IH. Diagnostic strategies for posttraumatic osteomyelitis: a survey amongst Dutch medical specialists demonstrates the need for a consensus protocol. *Eur J Trauma Emerg Surg* 2017;**44**:417–426. https://doi.org/10.1007/s00068-017-0783-9

119. Ayati N, Norouzi M, Sadeghi R, Erfani M, Gharedaghi M, Aryana K. Diagnostic value of 99mTc-ubiquicidin scintigraphy in differentiation between osteomyelitis and bone tumors. *Nucl Med Commun* 2017;**38**:885–90. https://doi.org/10.1097/MNM.0000000000000744

120. Ballani NS, Al-Huda FA, Khan HA, Al-Mohannadi S, Mahmood H, Al-Enezi F. The value of quantitative uptake of (99m)Tc-MDP and (99m)Tc-HMPAO white blood cells in detecting osteomyelitis in violated peripheral bones. *J Nucl Med Technol* 2007;**35**:91–5. https://doi.org/10.2967/jnmt.106.035402

121. Boronat-Ferrater M, Simó-Perdigó M, Cuberas-Borrós G, Aguadé-Bruix S, Dellepiane-Clarke F, Torrent-Llongarriu E, *et al.* Bone scintigraphy and radiolabeled white blood cell scintigraphy for the diagnosis of mandibular osteomyelitis. *Clin Nucl Med* 2011;**36**:273–6. https://doi.org/10.1097/RLU.0b013e31820a9ed5

122. Buhl T, Stentzer K, Hede A, Kjaer A, Hesse B. Bone infection in patients suspected of complicating osteomyelitis: the diagnostic value of dual isotope bone-granulocyte scintigraphy. *Clin Physiol Funct Imaging* 2005;**25**:20–6. https://doi.org/10.1111/j.1475-097X.2004.00581.x

123. Crerand S, Dolan M, Laing P, Bird M, Smith ML, Klenerman L. Diagnosis of osteomyelitis in neuropathic foot ulcers. *J Bone Joint Surg Br* 1996;**78**:51–5. https://doi.org/10.1302/0301-620X.78B1.0780051

124. Devillers A, Moisan A, Jean S, Arvieux C, Bourguet P. Technetium-99m hexamethylpropylene amine oxime leucocyte scintigraphy for the diagnosis of bone and joint infections: a retrospective study in 116 patients. *Eur J Nucl Med* 1995;**22**:302–7. https://doi.org/10.1007/BF00941845

125. Dillmann-Arroyo C, Cantú-Leal R, Campa-Núñez H, López-Cavazos C, Bermúdez-Argüelles M, Mejía-Herrera JC. [Application of the ubiquicidin 29-41 scan in the diagnosis of pyogenic vertebral osteomyelitis.] *Acta Ortop Mex* 2011;**25**:27–31.

126. Dutta P, Bhansali A, Mittal BR, Singh B, Masoodi SR. Instant 99mTc-ciprofloxacin scintigraphy for the diagnosis of osteomyelitis in the diabetic foot. *Foot Ankle Int* 2006;**27**:716–22. https://doi.org/10.1177/107110070602700911

127. el Esper I, Dacquet V, Paillard J, Bascoulergue G, Tahon MM, Fonroget J. 99Tcm-HMPAO-labelled leucocyte scintigraphy in suspected chronic osteomyelitis related to an orthopaedic device: clinical usefulness. *Nucl Med Commun* 1992;**13**:799–805. https://doi.org/10.1097/00006231-199211000-00005

128. Esterhai J, Alavi A, Mandell GA, Brown J. Sequential technetium-99m/gallium-67 scintigraphic evaluation of subclinical osteomyelitis complicating fracture nonunion. *J Orthop Res* 1985;**3**:219–25. https://doi.org/10.1002/jor.1100030212

129. Esterhai JL, Goll SR, McCarthy KE, Velchik M, Alavi A, Brighton CT, Heppenstall RB. Indium-111 leukocyte scintigraphic detection of subclinical osteomyelitis complicating delayed and nonunion long bone fractures: a prospective study. *J Orthop Res* 1987;**5**:1–6. https://doi.org/10.1002/jor.1100050102

130. Fernandez P, Monet A, Matei C, De Clermont H, Guyot M, Jeandot R, *et al.* 99mTc-HMPAO labelled white blood cell scintigraphy in patients with osteoarticular infection: the value of late images for diagnostic accuracy and interobserver reproducibility. *Eur J Clin Microbiol Infect Dis* 2008;**27**:1239–44. https://doi.org/10.1007/s10096-008-0563-x

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

131. Fihn SD, Larson EB, Nelp WB, Rudd TG, Gerber FH. Should single-phase radionuclide bone imaging be used in suspected osteomyelitis? *J Nucl Med* 1984;**25**:1080–8.

132. Glaudemans AW, de Vries EF, Vermeulen LE, Slart RH, Dierckx RA, Signore A. A large retrospective single-centre study to define the best image acquisition protocols and interpretation criteria for white blood cell scintigraphy with 99mTc-HMPAO-labelled leucocytes in musculoskeletal infections. *Eur J Nucl Med Mol Imaging* 2013;**40**:1760–9. https://doi.org/10.1007/s00259-013-2481-0

133. Govaert GAM, Bosch P, IJpma FFA, Glauche J, Jutte PC, Lemans JVC, *et al.* High diagnostic accuracy of white blood cell scintigraphy for fracture related infections: results of a large retrospective single-center study. *Injury* 2018;**49**:1085–90. https://doi.org/10.1016/j.injury.2018.03.018

134. Harvey J, Cohen MM. Technetium-99-labeled leukocytes in diagnosing diabetic osteomyelitis in the foot. *J Foot Ankle Surg* 1997;**36**:209–14. https://doi.org/10.1016/S1067-2516(97)80117-2

135. Harwood SJ, Camblin JG, Hakki S, Morrissey MA, Laven DL, Zangara LM, *et al.* Use of technetium antigranulocyte monoclonal antibody Fab' fragments for the detection of osteomyelitis. *Cell Biophys* 1994;**24–25**:99–107. https://doi.org/10.1007/BF02789220

136. Harwood SJ, Valdivia S, Hung GL, Quenzer RW. Use of Sulesomab, a radiolabeled antibody fragment, to detect osteomyelitis in diabetic patients with foot ulcers by leukoscintigraphy. *Clin Infect Dis* 1999;**28**:1200–5. https://doi.org/10.1086/514791

137. Herndon WA, Alexieva BT, Schwindt ML, Scott KN, Shaffer WO. Nuclear imaging for musculoskeletal infections in children. *J Pediatr Orthop* 1985;**5**:343–7. https://doi.org/10.1097/01241398-198505000-00017

138. Hotze AL, Briele B, Overbeck B, Kropp J, Gruenwald F, Mekkawy MA, *et al.* Technetium-99m-labeled anti-granulocyte antibodies in suspected bone infections. *J Nucl Med* 1992;**33**:526–31.

139. Jacobson AF, Harley JD, Lipsky BA, Pecoraro RE. Diagnosis of osteomyelitis in the presence of soft-tissue infection and radiologic evidence of osseous abnormalities: value of leukocyte scintigraphy. *AJR Am J Roentgenol* 1991;**157**:807–12. https://doi.org/10.2214/ajr.157.4.1892041

140. Johnson JE, Kennedy EJ, Shereff MJ, Patel NC, Collier BD. Prospective study of bone, indium-111-labeled white blood cell, and gallium-67 scanning for the evaluation of osteomyelitis in the diabetic foot. *Foot Ankle Int* 1996;**17**:10–16. https://doi.org/10.1177/107110079601700103

141. Kaim A, Maurer T, Ochsner P, Jundt G, Kirsch E, Mueller-Brand J. Chronic complicated osteomyelitis of the appendicular skeleton: diagnosis with technetium-99m labelled monoclonal antigranulocyte antibody-immunoscintigraphy. *Eur J Nucl Med* 1997;**24**:732–8. https://doi.org/10.1007/BF00879660

142. Kolindou A, Liu Y, Ozker K, Krasnow AZ, Isitman AT, Hellman RS, Collier BD. In-111 WBC imaging of osteomyelitis in patients with underlying bone scan abnormalities. *Clin Nucl Med* 1996;**21**:183–91. https://doi.org/10.1097/00003072-199603000-00001

143. Krznaric E, Roo MD, Verbruggen A, Stuyck J, Mortelmans L. Chronic osteomyelitis: diagnosis with technetium-99m-d, l-hexamethylpropylene amine oxime labelled leucocytes. *Eur J Nucl Med* 1996;**23**:792–7. https://doi.org/10.1007/BF00843708

144. Machens HG, Pallua N, Becker M, Mailaender P, Schaller E, Brenner P, *et al.* Technetium-99m human immunoglobulin (HIG): a new substance for scintigraphic detection of bone and joint infections. *Microsurgery* 1996;**17**:272–7. https://doi.org/10.1002/(SICI)1098-2752(1996)17:5<272::AID-MICR7>3.0.CO;2-L

145. Maurer AH, Millmond SH, Knight LC, Mesgarzadeh M, Siegel JA, Shuman CR, *et al.* Infection in diabetic osteoarthropathy: use of indium-labeled leukocytes for diagnosis. *Radiology* 1986;**161**:221–5. https://doi.org/10.1148/radiology.161.1.3763871

146. McCarthy K, Velchik MG, Alavi A, Mandell GA, Esterhai JL, Goll S. Indium-111-labeled white blood cells in the detection of osteomyelitis complicated by a pre-existing condition. *J Nucl Med* 1988;**29**:1015–21.

147. Melkun ET, Lewis VL. Evaluation of (111) indium-labeled autologous leukocyte scintigraphy for the diagnosis of chronic osteomyelitis in patients with grade IV pressure ulcers, as compared with a standard diagnostic protocol. *Ann Plast Surg* 2005;**54**:633–6. https://doi.org/10.1097/01.sap.0000164467.97551.ed

148. Mora Ríos FG, Isunza Ramírez A, López Marmolejo A, Palma Rosillo RM, Guízar Cuevas S, Mora Magaña I, Velázquez Pedroza VH. [Sensitivity and specificity of the Tc-99m ciprofloxacin scan in pediatric osteomyelitis.] *Acta Ortop Mex* 2010;**24**:248–51.

149. Nepola JV, Seabold JE, Marsh JL, Kirchner PT, el-Khoury GY. Diagnosis of infection in ununited fractures. Combined imaging with indium-111-labeled leukocytes and technetium-99m methylene diphosphonate. *J Bone Joint Surg Am* 1993;**75**:1816–22. https://doi.org/10.2106/00004623-199312000-00012

150. Nijhof MW, Oyen WJ, van Kampen A, Claessens RA, van der Meer JW, Corstens FH. Evaluation of infections of the locomotor system with indium-111-labeled human IgG scintigraphy. *J Nucl Med* 1997;**38**:1300–5.

151. Noriega-Álvarez E, Martínez Pimienta GA, Benítez Segura AM, Bajén Lázaro MT, Rodríguez-Gasén A, Rodríguez-Rubio Corona J, Mora-Salvadó J. Utility of 8 h and time decay-corrected acquisition scintigraphy with in-vitro labeled white blood cells for the diagnosis of osteoarticular infection. *Nucl Med Commun* 2017;**38**:500–8. https://doi.org/10.1097/MNM.0000000000000678

152. Oyen WJ, van Horn JR, Claessens RA, Slooff TJ, van der Meer JW, Corstens FH. Diagnosis of bone, joint, and joint prosthesis infections with In-111-labeled nonspecific human immunoglobulin G scintigraphy. *Radiology* 1992;**182**:195–9. https://doi.org/10.1148/radiology.182.1.1727281

153. Palestro CJ, Kim CK, Swyer AJ, Vallabhajosula S, Goldsmith SJ. Radionuclide diagnosis of vertebral osteomyelitis: indium-111-leukocyte and technetium-99m-methylene diphosphonate bone scintigraphy. *J Nucl Med* 1991;**32**:1861–5.

154. Palestro CJ, Kipper SL, Weiland FL, Love C, Tomas MB. Osteomyelitis: diagnosis with 99mTc-labeled antigranulocyte antibodies compared with diagnosis with ^{111}In-labeled leukocytes – initial experience. *Radiology* 2002;**223**:758–64. https://doi.org/10.1148/radiol.2233011072

155. Palestro CJ, Caprioli R, Love C, Richardson HL, Kipper SL, Weiland FL, Tomas MB. Rapid diagnosis of pedal osteomyelitis in diabetics with a technetium-99m-labeled monoclonal antigranulocyte antibody. *J Foot Ankle Surg* 2003;**42**:2–8. https://doi.org/10.1053/jfas.2003.50002

156. Sapienza MT, Hironaka F, Lima AL, Yamaga LY, Hamada E, Watanabe T, *et al.* [Evaluation of inflammatory activity in chronic osteomyelitis. Contribution of scintigraphy with polyclonal antibodies.] *Rev Assoc Med Bras* 2000;**46**:106–12. https://doi.org/10.1590/S0104-42302000000200004

157. Sayle BA, Fawcett HD, Yudt WM, Wang SC, Mader JT, Cierny G. Indium-111 chloride imaging with ununited fractures. *Clin Nucl Med* 1987;**12**:208–9. https://doi.org/10.1097/00003072-198703000-00010

158. Scheidler J, Leinsinger G, Pfahler M, Kirsch CM. Diagnosis of osteomyelitis. Accuracy and limitations of antigranulocyte antibody imaging compared to three-phase bone scan. *Clin Nucl Med* 1994;**19**:731–7. https://doi.org/10.1097/00003072-199408000-00019

159. Sciuk J, Brandau W, Vollet B, Stücker R, Erlemann R, Bartenstein P, *et al.* Comparison of technetium 99m polyclonal human immunoglobulin and technetium 99m monoclonal antibodies for imaging chronic osteomyelitis. First clinical results. *Eur J Nucl Med* 1991;**18**:401–7. https://doi.org/10.1007/BF02258431

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

160. Seabold JE, Nepola JV, Conrad GR, Marsh JL, Montgomery WJ, Bricker JA, Kirchner PT. Detection of osteomyelitis at fracture nonunion sites: comparison of two scintigraphic methods. *AJR Am J Roentgenol* 1989;**152**:1021–7. https://doi.org/10.2214/ajr.152.5.1021

161. Seabold JE, Nepola JV, Marsh JL, Hawes DR, Justin EP, Ponto JA, *et al.* Postoperative bone marrow alterations: potential pitfalls in the diagnosis of osteomyelitis with In-111-labeled leukocyte scintigraphy. *Radiology* 1991;**180**:741–7. https://doi.org/10.1148/radiology.180.3.1871288

162. Seabold JE, Ferlic RJ, Marsh JL, Nepola JV. Periarticular bone sites associated with traumatic injury: false-positive findings with In-111-labeled white blood cell and Tc-99m MDP scintigraphy. *Radiology* 1993;**186**:845–9. https://doi.org/10.1148/radiology.186.3.8430197

163. Seldin DW, Heiken JP, Feldman F, Alderson PO. Effect of soft-tissue pathology on detection of pedal osteomyelitis in diabetics. *J Nucl Med* 1985;**26**:988–93.

164. Singh B, Sunil HV, Sharma S, Prasad V, Kashyap R, Bhattacharya A, *et al.* Efficacy of indigenously developed single vial kit preparation of 99mTc-ciprofloxacin in the detection of bacterial infection: an Indian experience. *Nucl Med Commun* 2008;**29**:1123–9. https://doi.org/10.1097/MNM.0b013e328318b369

165. Sorsdahl OA, Goodhart GL, Williams HT, Hanna LJ, Rodriquez J. Quantitative bone gallium scintigraphy in osteomyelitis. *Skeletal Radiol* 1993;**22**:239–42. https://doi.org/10.1007/BF00197666

166. Tumeh SS, Aliabadi P, Seltzer SE, Weissman BN, McNeil BJ. Chronic osteomyelitis: the relative roles of scintigrams, plain radiographs, and transmission computed tomography. *Clin Nucl Med* 1988;**13**:710–15. https://doi.org/10.1097/00003072-198810000-00006

167. Unal SN, Birinci H, Baktiroğlu S, Cantez S. Comparison of Tc-99m methylene diphosphonate, Tc-99m human immune globulin, and Tc-99m-labeled white blood cell scintigraphy in the diabetic foot. *Clin Nucl Med* 2001;**26**:1016–21. https://doi.org/10.1097/00003072-200112000-00005

168. Vicente AG, Almoguera M, Alonso JC, Heffernan AJ, Gomez A, Contreras PI, *et al.* Diagnosis of orthopedic infection in clinical practice using Tc-99m sulesomab (antigranulocyte monoclonal antibody fragment Fab’2). *Clin Nucl Med* 2004;**29**:781–5. https://doi.org/10.1097/00003072-200412000-00001

169. Wukich DK, Abreu SH, Callaghan JJ, Van Nostrand D, Savory CG, Eggli DF, *et al.* Diagnosis of infection by preoperative scintigraphy with indium-labeled white blood cells. *J Bone Joint Surg Am* 1987;**69**:1353–60. https://doi.org/10.2106/00004623-198769090-00008

170. Wukich DK, Van Dam BE, Abreu SH. Preoperative indium-labeled white blood cell scintigraphy in suspected osteomyelitis of the axial skeleton. *Spine* 1988;**13**:1168–70. https://doi.org/10.1097/00007632-198810000-00020

171. Zeiger LS, Fox IM. Use of indium-111-labeled white blood cells in the diagnosis of diabetic foot infections. *J Foot Surg* 1990;**29**:46–51.

172. Martinez-Rios C, Jariwala M, Highmore K, Duffy KW, Laxer R, Stimec J. Sterile pyogenic arthritis pyoderma gangrenosum, and acne (PAPA) syndrome: Musculoskeletal imaging findings and common differential diagnosis in children. *Pediatr Radiol* 2016;**46**(Suppl. 1):1–372. https://doi.org/10.1007/s00247-016-3579-x

173. Wagner FW Jr. Treatment of the diabetic foot. *Compr Ther* 1984;**10**:29–38.

Appendix 1 Literature searches

The aim of the literature search was to systematically identify relevant studies of diagnostic imaging for the detection of osteomyelitis. The search strategy was developed by an information specialist with input from the review team and clinical advisors. The strategy was developed in MEDLINE (via Ovid) and included search terms for osteomyelitis and relevant diagnostic imaging techniques. No language, date, geographical or study design limits were applied. The MEDLINE strategy was adapted for use in the other resources searched.

The searches were carried out during August 2017 and updated in June 2018 to capture more recent studies. The following databases were searched: MEDLINE (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE), Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), Cumulative Index to Nursing and Allied Health (CINAHL) Plus, Database of Abstracts of Reviews of Effects (DARE), EMBASE, Health Technology Assessment (HTA) database and PubMed.

In addition, ClinicalTrials.gov and PROSPERO were searched for ongoing and unpublished studies. Relevant guidelines were identified through searches of the National Guidelines Clearing House, NHS Evidence, the NICE website and the Trip database.

The reference lists of relevant systematic reviews were manually checked to ensure that all relevant studies from previous reviews were identified. Abstracts from a number of conferences were consulted, including the Annual Meeting of the European Society of Paediatric Radiology, the Annual Meeting of the Society of Skeletal Radiology, the Annual Congress of the European Association of Nuclear Medicine and Paediatric Rheumatology. The search results were imported into EndNote X8 and de-duplicated. The complete search strategies can be found below.

Database search strategies

MEDLINE [Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)] via Ovid (http://ovidsp.ovid.com/)

Date range searched: 1946 to present.

Date searched: 9 August 2017.

Records retrieved: 10,184.

The search was updated on 20 June 2018 and identified 10,274 records.

Search strategy

1 Osteomyelitis/ [19,486]

2 Petrositis/ [74]

3 (osteomyelitis or osteomyelitides).ti,ab. [20,136]

4 ((bone$ or osseous or osteo) adj3 infect$).ti,ab. [6234]

5 ((spine or spines or spinal or vertebra$ or skeleton$ or skeletal or musculoskeletal) adj3 infect$).ti,ab. [4066]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

6 ((tibia or tibial or femur or humerus or humeral) adj3 infect$).ti,ab. [405]

7 majeed syndrome$.ti,ab. [32]

8 (petrositis or petrositides or Gradenigo$ or petrous apiciti$).ti,ab. [251]

9 or/[1-8 34,520]

10 exp Diagnostic Imaging/ [2,430,931]

11 (imag$ adj3 (diagnos$ or test$ or tool$ or procedure$ or method$ or technique$ or technolog$ or modalit$)).ti,ab. [168,689]

12 radiograph$.ti,ab. [191,546]

13 (x-ray$ or xray$ or roentgen$).ti,ab. [337,423]

14 (bone$ adj2 (scan$ or imag$)).ti,ab. [12,239]

15 (radionuclide adj2 (imag$ or scan$ or diagnos$)).ti,ab. [4856]

16 radioisotope$ scan$.ti,ab. [683]

17 (nuclear adj2 (medicine or imag$ or scan$)).ti,ab. [14,606]

18 ((magnetic resonance adj (imag$ or scan$ or tomograph$)) or MRI or MR imag$ or MR scan$ or MR tomograph$ or MRT or NMR or NMRI or fMRI or chemical shift imag$).ti,ab. [510,943]

19 ((compute$ adj2 tomograph$) or tomodensitometry or cine-CT).ti,ab. [253,513]

20 ((CT or CAT) adj (scan$ or imag$)).ti,ab. [105,989]

21 ((emission or positron or proton) adj2 tomograph$).ti,ab. [64,652]

22 (PET or PET-CT$ or PET?CT$ or CT-PET$ or CT?PET$).ti,ab. [79,016]

23 (SPECT or SPECT-CT$ or SPECT?CT$ or CT-SPECT$ or CT?SPECT$).ti,ab. [25,475]

24 (SPET or SPET-CT$ or SPET?CT$ or CT-SPET$ or CT?SPET$).ti,ab. [1321]

25 (PET-MRI$ or PET?MRI$).ti,ab. [1227]

26 Fluorodeoxyglucose F18/ [25,483]

27 ((FDG or fluorodeoxyglucose) adj4 (imag$ or scan$)).ti,ab. [9778]

28 (FDG-PET$ or FDG?PET$).ti,ab. [20,415]

29 (ultrasound$ or ultrasonograph$ or echograph$ or ultrasonic or sonograph$ or echotomograph$ or echogram$ or echoscop$ or echosound$).ti,ab. [351,485]

30 (scintigraph$ or scintiphotograph$ or scintigram$).ti,ab. [47,128]

31 (scintiscan$ or immunoscintigra$ or leu#oscintigra$).ti,ab. [2486]

32 ((leu#ocyte$ adj2 (scan$ or imag$)) or leu#oscan$).ti,ab. [697]

33 (white blood cell$ adj2 (scan$ or imag$)).ti,ab. [195]

34 (WBC scan$ or WBCS).ti,ab. [2034]

35 or/[10-34 3,217,417]

36 9 and [35 10,908]

37 exp animals/not humans/ [4,448,945]

38 36 not [37 10,184]

Key

/ = indexing term (MeSH heading).

exp = exploded indexing term (MeSH heading).

$ = truncation.

? = optional wildcard, stands for zero or one character.

# = mandated wildcard, stands for one character within a word.

ti,ab = terms in either title or abstract fields.

adj3 = terms within three words of each other (any order).

Cochrane Central Register of Controlled Trials via Wiley (http://onlinelibrary.wiley.com/)

Date range searched: issue 7 of 12, July 2017.

Date searched: 9 August 2017.

Records retrieved: 147.

The strategy below was used to search CENTRAL and CDSR.

The search was updated on 20 June 2018 and identified 167 records from CENTRAL.

Search strategy

#1 MeSH descriptor: [Osteomyelitis] this term only [131]

#2 MeSH descriptor: [Petrositis] this term only [0]

#3 (osteomyelitis or osteomyelitides):ti,ab,kw [367]

#4 ((bone* or osseous or osteo) near/3 infect*):ti,ab,kw [307]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

#5 ((spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal) near/3 infect*):ti,ab,kw [184]

#6 ((tibia or tibial or femur or humerus or humeral) near/3 infect*):ti,ab,kw [55]

#7 (majeed next syndrome*):ti,ab,kw [0]

#8 (petrositis or petrositides or Gradenigo* or petrous next apiciti*):ti,ab,kw [3]

#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 [845]

#10 MeSH descriptor: [Diagnostic Imaging] explode all trees [45,280]

#11 (imag* near/3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*)):ti,ab,kw [10,201]

#12 radiograph*:ti,ab,kw [14,487]

#13 (x-ray* or xray* or roentgen*):ti,ab,kw [12,891]

#14 (bone* near/2 (scan* or imag*)):ti,ab,kw [513]

#15 (radionuclide near/2 (imag* or scan* or diagnos*)):ti,ab,kw [1683]

#16 (radioisotope* next scan*):ti,ab,kw [6]

#17 (nuclear near/2 (medicine or imag* or scan*)):ti,ab,kw [1197]

#18 ((magnetic next resonance next (imag* or scan* or tomograph*)) or MRI or MR next imag* or MR next scan* or MR next tomograph* or MRT or NMR or NMRI or fMRI or chemical next shift next imag*):ti,ab,kw [18,678]

#19 ((compute* near/2 tomograph*) or tomodensitometry or cine-CT):ti,ab,kw [13,251]

#20 ((CT or CAT) next (scan* or imag*)):ti,ab,kw [3595]

#21 ((emission or positron or proton) near/2 tomograph*):ti,ab,kw [5445]

#22 (PET or PET-CT* or PET*CT* or CT-PET* or CT*PET*):ti,ab,kw [3930]

#23 (SPECT or SPECT-CT* or SPECT*CT* or CT-SPECT* or CT*SPECT*):ti,ab,kw [1385]

#24 (SPET or SPET-CT* or SPET*CT* or CT-SPET* or CT*SPET*):ti,ab,kw [96]

#25 (PET-MRI* or PET*MRI*):ti,ab,kw [49]

#26 MeSH descriptor: [Fluorodeoxyglucose F18] this term only [815]

#27 ((FDG or fluorodeoxyglucose) near/4 (imag* or scan*)):ti,ab,kw [602]

#28 (FDG-PET* or FDG*PET*):ti,ab,kw [1142]

#29 (ultrasound* or ultrasonograph* or echograph* or ultrasonic or sonograph* or echotomograph* or echogram* or echoscop* or echosound*):ti,ab,kw [24,819]

#30 (scintigraph* or scintiphotograph* or scintigram*):ti,ab,kw [2121]

#31 (scintiscan* or immunoscintigra* or leu?oscintigra*):ti,ab,kw [868]

#32 ((leu?ocyte* near/2 (scan* or imag*)) or leu?oscan$):ti,ab,kw [29]

#33 ("white blood cell" near/2 (scan* or imag*)):ti,ab,kw [5]

#34 ("white blood cells" near/2 (scan* or imag*)):ti,ab,kw [0]

#35 (WBC next scan* or WBCS):ti,ab,kw [136]

#36 {or #10 to #35} [93,342]

#37 #9 and #36 [164]

#38 #9 and #36 in Cochrane Reviews (Reviews and Protocols) 6

#39 #9 and #36 in Trials [147]

Key

MeSH descriptor = indexing term (MeSH heading).

* = truncation.

? = wildcard, matches a single character within a word.

ti,ab,kw = terms in either title or abstract or keyword fields.

near/3 = terms within three words of each other (any order).

next = terms next to each other (order specified).

" " = phrase search.

Cochrane Database of Systematic Reviews via Wiley (http://onlinelibrary.wiley.com/)

Date range searched: issue 8 of 12, August 2017.

Date searched: 9 August 2017.

Records retrieved: 6.

See above under CENTRAL for search strategy used.

The search was updated on 20 June 2018 and identified six records from CDSR.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Cumulative Index to Nursing and Allied Health Plus via EBSCO (www.ebscohost.com/)

Date range searched: inception to 8 August 2017.

Date searched: 9 August 2017.

Records retrieved: 1590.

The search was updated on 20 June 2018 and identified 1781 records.

Search strategy

S1 (MH "Osteomyelitis") [2641]

S2 (MH "Gradenigo Syndrome") [15]

S3 TI (osteomyelitis or osteomyelitides) OR AB (osteomyelitis or osteomyelitides) [2690]

S4 TI ((bone* or osseous or osteo) N3 infect*)) OR AB ((bone* or osseous or osteo) N3 infect*)) [906]

S5 TI ((spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal) N3 infect*) OR AB ((spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal) N3 infect*) [1003]

S6 TI ((tibia or tibial or femur or humerus or humeral) N3 infect*) OR AB ((tibia or tibial or femur or humerus or humeral) N3 infect*) [153]

S7 TI majeed N1 syndrome* OR AB majeed N1 syndrome* [12]

S8 TI (petrositis or petrositides or Gradenigo* or petrous N1 apiciti*) OR AB (petrositis or petrositides or Gradenigo* or petrous N1 apiciti*) [32]

S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 [5277]

S10 (MH "Diagnostic Imaging+") [292,403]

S11 TI (imag* N3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*)) OR AB (imag* N3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*)) [28,469]

S12 TI radiograph* OR AB radiograph* [38,222]

S13 TI (x-ray* or xray* or roentgen*) OR AB (x-ray* or xray* or roentgen*) [16,289]

S14 TI ((bone* N2 (scan* or imag*)) OR AB ((bone* N2 (scan* or imag*)) [1884]

S15 TI (radionuclide N2 (imag* or scan* or diagnos*)) OR AB (radionuclide N2 (imag* or scan* or diagnos*)) [394]

S16 TI radioisotope* N1 scan* OR AB radioisotope* N1 scan* [26]

S17 TI (nuclear N2 (medicine or imag* or scan*)) OR AB (nuclear N2 (medicine or imag* or scan*)) [1840]

S18 TI (("magnetic resonance" or MR) N1 (imag* or scan* or tomograph*)) OR AB (("magnetic resonance" or MR) N1 (imag* or scan* or tomograph*)) [39,205]

S19 TI (MRI or MRT or NMR or NMRI or fMRI) OR AB (MRI or MRT or NMR or NMRI or fMRI) [39,230]

S20 TI chemical N1 shift N1 imag* OR AB chemical N1 shift N1 imag* [71]

S21 TI ((compute* N2 tomograph*) or tomodensitometry or cine-CT) OR AB ((compute* N2 tomograph*) or tomodensitometry or cine-CT) [42,942]

S22 TI (CAT N1 (scan* or imag*)) OR AB (CAT N1 (scan* or imag*)) [100]

S23 TI ((emission or positron or proton) N2 tomograph*) OR AB ((emission or positron or proton) N2 tomograph*) [9054]

S24 TI (PET or PET-CT* or PET#CT* or CT-PET* or CT#PET*) OR AB (PET or PET-CT* or PET#CT* or CT-PET* or CT#PET*) [59,312]

S25 TI (SPECT or SPECT-CT* or SPECT#CT* or CT-SPECT* or CT#SPECT*) OR AB (SPECT or SPECT-CT* or SPECT#CT* or CT-SPECT* or CT#SPECT*) [54,720]

S26 TI (SPET or SPET-CT* or SPET#CT* or CT-SPET* or CT#SPET*) OR AB (SPET or SPET-CT* or SPET#CT* or CT-SPET* or CT#SPET*) [52,699]

S27 TI (PET-MRI* or PET#MRI*) OR AB (PET-MRI* or PET#MRI*) [235]

S28 (MH "Fludeoxyglucose F 18") [4313]

S29 TI ((FDG or fluorodeoxyglucose) N4 (imag* or scan*)) OR AB ((FDG or fluorodeoxyglucose) N4 (imag* or scan*)) [1791]

S30 TI (FDG-PET* or FDG#PET*) OR AB (FDG-PET* or FDG#PET*) [3659]

S31 TI (ultrasound* or ultrasonograph* or echograph* or ultrasonic or sonograph* or echotomograph* or echogram* or echoscop* or echosound*) OR AB (ultrasound* or ultrasonograph* or echograph* or ultrasonic or sonograph* or echotomograph* or echogram* or echoscop* or echosound*) [54,423]

S32 TI (scintigraph* or scintiphotograph* or scintigram*) OR AB (scintigraph* or scintiphotograph* or scintigram*) [2711]

S33 TI (scintiscan* or immunoscintigra* or leu?oscintigra*) OR AB (scintiscan* or immunoscintigra* or leu?oscintigra*) [34]

S34 TI (leu?ocyte* N2 (scan* or imag*)) OR AB (leu?ocyte* N2 (scan* or imag*)) [43]

S35 TI leu?oscan* OR AB leu?oscan* [1]

S36 TI ("white blood cell" N2 (scan* or imag*)) OR AB ("white blood cell" N2 (scan* or imag*)) [16]

S37 TI ("white blood cells" N2 (scan* or imag*)) OR AB ("white blood cells" N2 (scan* or imag*)) [4]

S38 TI ((WBC N1 scan*) or WBCS) OR AB ((WBC N1 scan*) or WBCS) [1459]

S39 S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 [405,211]

S40 S9 AND S39 [1590]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Key

MH = indexing term (CINAHL heading).

* = truncation.

# = wildcard, stands for zero or one character.

? = wildcard, stands for one character.

TI = terms in the title.

AB = terms in the abstract.

N3 = terms within three words of each other (any order).

Database of Abstracts of Reviews of Effects via (www.crd.york.ac.uk/CRDWeb/)

Date range searched: inception to 31 March 2015.

Date searched: 9 August 2017.

Records retrieved: 14.

The strategy below was used to search DARE and the HTA database.

The search of the DARE database was not updated as it closed at the end of March 2015.

Search strategy

1 MeSH DESCRIPTOR Osteomyelitis IN DARE,HTA [22]

2 MeSH DESCRIPTOR Petrositis IN DARE,HTA [0]

3 (osteomyelitis or osteomyelitides) IN DARE, HTA [51]

4 ((bone* or osseous or osteo) NEAR3 infect*) OR (infect* NEAR3 (bone* or osseous or osteo)) IN DARE, HTA [21]

5 ((spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal) NEAR3 infect*) OR (infect* NEAR3 (spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal)) IN DARE, HTA [22]

6 ((tibia or tibial or femur or humerus or humeral) NEAR3 infect*) OR (infect* NEAR3 (tibia or tibial or femur or humerus or humeral)) IN DARE, HTA [2]

7 (majeed syndrome*) IN DARE, HTA [0]

8 (petrositis or petrositides or Gradenigo* or petrous apiciti*) IN DARE, HTA [0]

9 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 [88]

10 MeSH DESCRIPTOR Diagnostic imaging EXPLODE ALL TREES IN DARE,HTA [2969]

11 (imag* NEAR3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*)) OR ((diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*) NEAR3 imag*) IN DARE, HTA [1173]

12 (radiograph*) IN DARE, HTA [1092]

13 (x-ray* or xray* or roentgen*) IN DARE, HTA [873]

14 (bone* NEAR2 (scan* or imag*)) OR ((scan* or imag*) NEAR2 bone*) IN DARE, HTA [48]

15 (radionuclide NEAR2 (imag* or scan* or diagnos*)) OR ((imag* or scan* or diagnos*) NEAR2 radionuclide) IN DARE, HTA [313]

16 (radioisotope* scan*) IN DARE, HTA [1]

17 (nuclear NEAR2 (medicine or imag* or scan*)) OR ((medicine or imag* or scan*) NEAR2 nuclear) IN DARE, HTA [102]

18 (MRI or MRT or NMR or NMRI or fMRI) IN DARE, HTA [506]

19 (chemical shift imag*) IN DARE, HTA [1]

20 (compute* NEAR2 tomograph*) IN DARE, HTA [942]

21 (tomograph* NEAR2 compute*) IN DARE, HTA [809]

22 (tomodensitometry or cine-CT) IN DARE, HTA [0]

23 (CT scan* or CAT scan*) OR (CT imag* or CT imag*) IN DARE, HTA [186]

24 ((emission or positron or proton) NEAR2 tomograph*) OR (tomograph* NEAR2 (emission or positron or proton)) IN DARE, HTA [602]

25 (PET or PET-CT* or PET/CT* or PETCT* or CT-PET* or CT/PET* or CTPET*) IN DARE, HTA [437]

26 (SPECT or SPECT-CT* or SPECT/CT* or SPECTCT* or CT-SPECT* or CT/SPECT* or CTSPECT*) IN DARE, HTA [73]

27 (SPET or SPET-CT* or SPET/CT* or SPETCT* or CT-SPET* or CT/SPET* or CTSPET*) IN DARE, HTA [2]

28 (PET-MRI* or PET/MRI* or PETMRI*) IN DARE, HTA [8]

29 MeSH DESCRIPTOR Fluorodeoxyglucose F18 IN DARE,HTA [223]

30 ((FDG or fluorodeoxyglucose) NEAR4 (imag* or scan*)) OR ((imag* or scan*) NEAR4 (FDG or fluorodeoxyglucose)) IN DARE, HTA [133]

31 (FDG-PET* or FDG/PET* or FDGPET*) IN DARE, HTA [200]

32 (ultrasound* or ultrasonograph* or echograph* or ultrasonic or sonograph* or echotomograph* or echogram* or echoscop* or echosound*) IN DARE, HTA [1722]

33 (scintigraph* or scintiphotograph* or scintigram*) IN DARE, HTA [105]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

34 (scintiscan* or immunoscintigra* or leucoscintigra* or leukoscintigra*) IN DARE, HTA [1]

35 ((leucocyte* or leukocyte*) NEAR2 (scan* or imag*)) OR ((scan* or imag*) NEAR2 (leucocyte* or leukocyte*)) IN DARE, HTA [2]

36 (leukoscan* or leucoscan*) IN DARE, HTA [1]

37 (white blood cell* NEAR2 (scan* or imag*)) OR ((scan* or imag*) NEAR2 white blood cell*) IN DARE, HTA [1]

38 (WBC scan* or WBCS) IN DARE, HTA [4]

39 (magnetic resonance imag* or magnetic resonance scan* or magnetic resonance tomograph*) IN DARE, HTA [852]

40 (MR imag* or MR scan* or MR tomograph*) IN DARE, HTA [31]

41 #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 [5045]

42 #9 AND #41 (16) (14 DARE, 2 HTA)

Key

MeSH DESCRIPTOR = indexing term (MeSH heading).

* = truncation.

NEAR3 = terms within three words of each other (order specified).

EMBASE via Ovid (http://ovidsp.ovid.com/)

Date range searched: 1974 to 8 August 2017.

Date searched: 9 August 2017.

Records retrieved: 14,494.

The search was updated on 20 June 2018 and identified 15,225 records.

Search strategy

1 osteomyelitis/or chronic osteomyelitis/or hematogenous osteomyelitis/or majeed syndrome/or petrositis/ [29,324]

2 (osteomyelitis or osteomyelitides).ti,ab. [22,929]

3 ((bone$ or osseous or osteo) adj3 infect$).ti,ab. [7658]

4 ((spine or spines or spinal or vertebra$ or skeleton$ or skeletal or musculoskeletal) adj3 infect$).ti,ab. [5024]

5 ((tibia or tibial or femur or humerus or humeral) adj3 infect$).ti,ab. [476]

6 majeed syndrome$.ti,ab. [46]

7 (petrositis or petrositides or Gradenigo$ or petrous apiciti$).ti,ab. [263]

8 or/1-7 [44,156]

9 exp diagnostic imaging/ [143,551]

10 (imag$ adj3 (diagnos$ or test$ or tool$ or procedure$ or method$ or technique$ or technolog$ or modalit$)).ti,ab. [232,999]

11 radiography/ [253,398]

12 exp bone radiography/ [62,582]

13 X ray/ [59,947]

14 radiograph$.ti,ab. [227,261]

15 (x-ray$ or xray$ or roentgen$).ti,ab. [365,257]

16 (bone$ adj2 (scan$ or imag$)).ti,ab. [17,134]

17 (radionuclide adj2 (imag$ or scan$ or diagnos$)).ti,ab. [5946]

18 radioisotope$ scan$.ti,ab. [758]

19 (nuclear adj2 (medicine or imag$ or scan$)).ti,ab. [26,429]

20 exp nuclear magnetic resonance imaging/ [759,296]

21 ((magnetic resonance adj (imag$ or scan$ or tomograph$)) or MRI or MR imag$ or MR scan$ or MR tomograph$ or MRT or NMR or NMRI or fMRI or chemical shift imag$).ti,ab. [681,481]

22 exp computer assisted tomography/ [846,193]

23 ((compute$ adj2 tomograph$) or tomodensitometry or cine-CT).ti,ab. [302,150]

24 ((CT or CAT) adj (scan$ or imag$)).ti,ab. [164,664]

25 exp x-ray tomography/ [11,890]

26 exp emission tomography/ [146,279]

27 ((emission or positron or proton) adj2 tomograph$).ti,ab. [79,931]

28 (PET or PET-CT$ or PET?CT$ or CT-PET$ or CT?PET$).ti,ab. [128,903]

29 (SPECT or SPECT-CT$ or SPECT?CT$ or CT-SPECT$ or CT?SPECT$).ti,ab. [39,774]

30 (SPET or SPET-CT$ or SPET?CT$ or CT-SPET$ or CT?SPET$).ti,ab. [1533]

31 (PET-MRI$ or PET?MRI$).ti,ab. [2339]

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

32 fluorodeoxyglucose f 18/ [48,488]

33 ((FDG or fluorodeoxyglucose) adj4 (imag$ or scan$)).ti,ab. [17,045]

34 (FDG-PET$ or FDG?PET$).ti,ab. [34,822]

35 exp echography/ [640,877]

36 (ultrasound$ or ultrasonograph$ or echograph$ or ultrasonic or sonograph$ or echotomograph$ or echogram$ or echoscop$ or echosound$).ti,ab. [479,205]

37 exp scintiscanning/ [172,675]

38 (scintigraph$ or scintiphotograph$ or scintigram$).ti,ab. [62,147]

39 (scintiscan$ or immunoscintigra$ or leu#oscintigra$).ti,ab. [3034]

40 ((leu#ocyte$ adj2 (scan$ or imag$)) or leu#oscan$).ti,ab. [857]

41 (white blood cell$ adj2 (scan$ or imag$)).ti,ab. [228]

42 (WBC scan$ or WBCS).ti,ab. [3270]

43 or/9-42 [3,184,438]

44 8 and 43 [15,191]

45 (animal/or animal experiment/or animal model/or animal tissue/or nonhuman/) not exp human/ [5,733,000]

46 44 not 45 [14,494]

Key

/ = indexing term (MeSH heading).

exp = exploded indexing term (MeSH heading).

$ = truncation.

? = optional wildcard, stands for zero or one character.

# = mandated wildcard, stands for one character within a word.

ti,ab = terms in either title or abstract fields.

adj3 = terms within three words of each other (any order).

Health Technology Assessment database via (http://www.crd.york.ac.uk/CRDWeb/)

Date range searched: inception to 8 August 2017.

Date searched: 9 August 2017.

Records retrieved: 2.

See above under DARE for search strategy used.

The search of the HTA database was updated on 21 June 2018 and identified two records.

PubMed (http://www.ncbi.nlm.nih.gov/pubmed/)

Date range searched: inception to 9 August 2017.

Date searched: 9 August 2017.

Records retrieved: 1534.

The search was updated on 20 June 2018 and identified 1908 records.

Search strategy

Search (((((((((((((“Osteomyelitis”[Mesh:noexp]) OR “Petrositis”[Mesh:noexp]) OR ((osteomyelitis[Title/Abstract] OR osteomyelitides[Title/Abstract]))) OR (((bone*[Title/Abstract] OR osseous[Title/Abstract] OR osteo[Title/Abstract])) AND infect*[Title/Abstract])) OR (((spine[Title/Abstract] OR spines[Title/Abstract] OR spinal[Title/Abstract] OR vertebra*[Title/Abstract] OR skeleton*[Title/Abstract] OR skeletal[Title/Abstract] OR musculoskeletal [Title/Abstract])) AND infect*[Title/Abstract])) OR (((tibia[Title/Abstract] OR tibial[Title/Abstract] OR femur[Title/Abstract] OR humerus[Title/Abstract] OR humeral[Title/Abstract])) AND infect*[Title/Abstract])) OR ((majeed syndrome[Title/Abstract]) OR majeed syndromes[Title/Abstract])) OR ((petrositis[Title/Abstract] OR petrositides [Title/Abstract] OR Gradenigo$[Title/Abstract] OR “petrous apicitis”[Title/Abstract] OR “petrous apicitides”[Title/Abstract])))) AND ((((((((((((((((((((((((((((((((((“Diagnostic Imaging”[Mesh]) OR ((((diagnos*[Title/Abstract] OR test*[Title/Abstract] OR tool*[Title/Abstract] OR procedure*[Title/Abstract] OR method*[Title/Abstract] OR technique* [Title/Abstract] OR technolog*[Title/Abstract] OR modalit*[Title/Abstract])) AND imag*[Title/Abstract])) OR radiograph*[Title/Abstract]) OR ((x-ray*[Title/Abstract] OR xray*[Title/Abstract] OR roentgen*[Title/Abstract]))) OR ((bone*[Title/Abstract]) AND (scan*[Title/Abstract] OR imag*[Title/Abstract]))) OR ((radionuclide[Title/Abstract]) AND (imag*[Title/Abstract] OR scan*[Title/Abstract] OR diagnos*[Title/Abstract]))) OR radioisotope scan*[Title/Abstract]) OR ((nuclear[Title/Abstract]) AND (medicine[Title/Abstract] OR imag*[Title/Abstract] OR scan*[Title/Abstract]))) OR ((((((((magnetic resonance imag*[Title/Abstract]) OR magnetic resonance scan*[Title/Abstract]) OR magnetic resonance tomograph*[Title/Abstract]) OR MR imag*[Title/Abstract]) OR MR scan*[Title/Abstract]) OR MR tomograph*[Title/Abstract]) OR (MRI[Title/Abstract] OR MRT[Title/Abstract] OR NMR[Title/Abstract] OR NMRI [Title/Abstract] OR fMRI[Title/Abstract])) OR chemical shift imag*[Title/Abstract])) OR ((compute*[Title/Abstract]) AND tomograph*[Title/Abstract])) OR ((tomodensitometry[Title/Abstract] OR cine-CT[Title/Abstract]))) OR (((CT scan*[Title/Abstract] OR CT imag*[Title/Abstract])) OR (CAT scan*[Title/Abstract] OR CAT imag*[Title/Abstract]))) OR (((emission[Title/Abstract] OR positron[Title/Abstract] OR proton[Title/Abstract])) AND tomograph*[Title/Abstract])) OR ((PET[Title/Abstract] OR PET-CT*[Title/Abstract] OR PET/CT*[Title/Abstract] OR PETCT*[Title/Abstract] OR CT-PET*[Title/Abstract] OR CT/PET$[Title/Abstract] OR CTPET*[Title/Abstract]))) OR ((SPECT [Title/Abstract] OR SPECT-CT*[Title/Abstract] OR SPECT/CT*[Title/Abstract] OR SPECTCT*[Title/Abstract] OR CT-SPECT*[Title/Abstract] OR CT/SPECT*[Title/Abstract] OR CTSPECT*[Title/Abstract]))) OR ((SPET[Title/Abstract] OR SPET-CT*[Title/Abstract] OR SPET/CT*[Title/Abstract] OR SPETCT*[Title/Abstract] OR CT-SPET* [Title/Abstract] OR CT/SPET*[Title/Abstract] OR CTSPET*[Title/Abstract]))) OR ((PET-MRI*[Title/Abstract] OR PET/MRI*[Title/Abstract] OR PETMRI*[Title/Abstract]))) OR “Fluorodeoxyglucose F18”[Mesh:noexp]) OR ((FDG-PET* [Title/Abstract] OR FDG/PET*[Title/Abstract] OR FDGPET*[Title/Abstract]))) OR ((ultrasound*[Title/Abstract] OR ultrasonograph*[Title/Abstract] OR echograph*[Title/Abstract] OR ultrasonic[Title/Abstract] OR sonograph* [Title/Abstract] OR echotomograph*[Title/Abstract] OR echogram*[Title/Abstract] OR echoscop*[Title/Abstract] OR echosound*[Title/Abstract]))) OR ((scintigraph*[Title/Abstract] OR scintiphotograph*[Title/Abstract] OR scintigram*[Title/Abstract]))) OR ((scintiscan*[Title/Abstract] OR immunoscintigra*[Title/Abstract] OR leucoscintigra*[Title/Abstract] OR leukoscintigra*[Title/Abstract]))) OR (((leucocyte*[Title/Abstract] OR

© Queen’s Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

leukocyte*[Title/Abstract])) AND (scan*[Title/Abstract] OR imag*[Title/Abstract]))) OR ((leukoscan*[Title/Abstract] OR leucoscan*[Title/Abstract]))) OR ((white blood cell*[Title/Abstract]) AND (scan*[Title/Abstract] OR imag*[Title/Abstract]))) OR ((WBC scan*[Title/Abstract] OR WBCS[Title/Abstract]))) OR ((((FDG[Title/Abstract] OR fluorodeoxyglucose[Title/Abstract])) AND (imag*[Title/Abstract] OR scan*[Title/Abstract])))))) NOT ((animals [mh] NOT humans[mh])))) AND ((pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb]))

The above search strategy incorporates the following search line to limit to studies found in PubMed but not available in Ovid MEDLINE: (pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb]).[172]

Key

[Mesh] = exploded indexing term (MeSH heading).

[Mesh:noexp] = indexing term (MeSH heading) not exploded.

* = truncation.

[Title/Abstract]) = terms in either title or abstract fields.

" " = phrase search.

Ongoing studies search strategies

ClinicalTrials.gov (https://clinicaltrials.gov/)

Date range searched: inception to 9 August 2017.

Searched date: 9 August 2017.

Records retrieved: 66.

The search was updated on 20 June 2018 and identified 76 records.

Searched in the condition field: osteomyelitis OR "bone infection".

PROSPERO (www.crd.york.ac.uk/PROSPERO/)

Date range searched: inception to 9 August 2017.

Searched date: 9 August 2017.

Records retrieved: 20.

The search was updated on 21 June 2018 and identified 29 records.

Search strategy

#1 MeSH DESCRIPTOR osteomyelitis [6]

#2 MeSH DESCRIPTOR Petrositis [0]

#3 osteomyelitis OR osteomyelitides [38]

#4 (bone* or osseous or osteo) adj3 infect* [18]

#5 (spine or spines or spinal or vertebra* or skeleton* or skeletal or musculoskeletal) adj3 infect* [29]

#6 (tibia or tibial or femur or humerus or humeral) adj3 infect* [2]

#7 majeed syndrome* [0]

#8 petrositis or petrositides or Gradenigo* or petrous apiciti* [0]

#9 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 [81]

#10 MeSH DESCRIPTOR Diagnostic Imaging EXPLODE ALL TREES [637]

#11 imag* adj3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*) [460]

#12 imag* ADJ3 (diagnos* or test* or tool* or procedure* or method* or technique* or technolog* or modalit*) [460]

#13 radiograph* [556]

#14 x-ray* or xray* or roentgen* [313]

#15 bone* adj2 (scan* or imag*) [16]

#16 radionuclide adj2 (imag* or scan* or diagnos*) [6]

#17 radioisotope* scan* [0]

#18 nuclear adj2 (medicine or imag* or scan*) [31]

#19 “magnetic resonance” adj (imag* or scan* or tomograph*) [373]

#20 “magnetic resonance” adj1 (imag* or scan* or tomograph*) [358]

#21 “MR” adj (imag* or scan* or tomograph*) [45}

#22 MRI or MRT or NMR or NMRI or fMRI or chemical shift imag* [678]

#23 compute* adj2 tomograph* [383]

#24 tomodensitometry or cine-CT [0]

#25 (CT or CAT) adj (scan* or imag*) [260]

#26 (emission or positron or proton) adj2 tomograph* [76]

#27 PET or PET-CT* or PET*CT* or CT-PET* or CT*PET* [142]

#28 SPECT or SPECT-CT* or SPECT*CT* or CT-SPECT* or CT*SPECT* [44]

#29 SPET or SPET-CT* or SPET*CT* or CT-SPET* or CT*SPET* [0]

© Queen’s Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

#30 PET-MRI* or PET*MRI* [5]

#31 MeSH DESCRIPTOR Fluorodeoxyglucose F18 [15]

#32 (FDG or fluorodeoxyglucose) adj4 (imag* or scan*) [13]

#33 FDG-PET* or FDG*PET* [24]

#34 ultrasound* or ultrasonograph* or echograph* or ultrasonic or sonograph* or echotomograph* or echogram* or echoscop* or echosound* [838]

#35 scintigraph* or scintiphotograph* or scintigram* [28]

#36 scintiscan* or immunoscintigra* or leucoscintigra* or leukoscintigra* [0]

#37 (leucocyte* or leukocyte*) adj2 (scan* or imag*) [0]

#38 white blood cell* adj2 (scan* or imag*) [0]

#39 WBC scan* or WBCS [4]

#40 #39 OR #38 OR #37 OR #36 OR #35 OR #34 OR #33 OR #32 OR #31 OR #30 OR #29 OR #28 OR #27 OR #26 OR #25 OR #24 OR #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 [2500]

#41 #40 AND #9 [20]

#42 leukoscan* or leucoscan* [0]

#43 #40 OR #42 [2500]

#44 #9 AND #43 [20]

Key

MeSH DESCRIPTOR = indexing term (MeSH heading).

* = truncation.

ADJ3 = terms within three words of each other.

Guideline searches

National Guideline Clearinghouse (www.guideline.gov/)

Date searched: 31 August 2017.

Searched for osteomyelitis – 38 results.

The search was updated on 21 June 2018 and identified 33 results.

NHS Evidence (www.evidence.nhs.uk/)

Date searched: 31 August 2017.

Searched for osteomyelitis, filtered results to limit to guidance – 161 results.

The search was updated on 21 June 2018 and identified 141 results.

National Institute for Health and Care Excellence (www.nice.org.uk/guidance)

Date searched: 31 August 2017.

Browsed the diagnostics guidance page (www.nice.org.uk/guidance/published?type=dg) – zero relevant records found.

Searched for the term osteomyelitis in the general website search box – 21 results browsed – two relevant records found.

The search was updated on 21 June 2018 but did not identify any further relevant records.

Trip (www.tripdatabase.com)

Date searched: 31 August 2017.

Searched for osteomyelitis – 95 results – browsed for relevance and checked against previous search results in EndNote for duplicates. One relevant record identified.

The search was updated on 21 June 2018 but did not identify any further relevant records.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 2 Results of critical appraisal of diagnostic accuracy studies (QUADAS-2)

TABLE 15 The QUADAS-2 signalling questions for risk-of-bias assessment

Domain			
Domain 1: patient selection	**Domain 2: index test**	**Domain 3: reference standard**	**Domain 4: flow and timing**
SQ 1: was a consecutive or random sample of patients enrolled?	SQ 1: were the index test results interpreted without knowledge of the results of the reference standard?	SQ 1: is the reference standard likely to correctly classify the target condition?	SQ 1: was there an appropriate interval between the index test(s) and reference standard?
SQ 2: was a case–control design avoided?	SQ 2: if a threshold was used, was it prespecified?	SQ 2: were the reference standard results interpreted without knowledge of the results of the index test?	SQ 2: did all patients receive the same reference standard?
SQ 3: did the study avoid inappropriate exclusions?	SQ 3: (optional: multiple tests) were the index test results interpreted without knowledge of the results of other index tests?	SQ3: not applicable	SQ 3: were all patients included in the analysis?

SQ, signalling question.

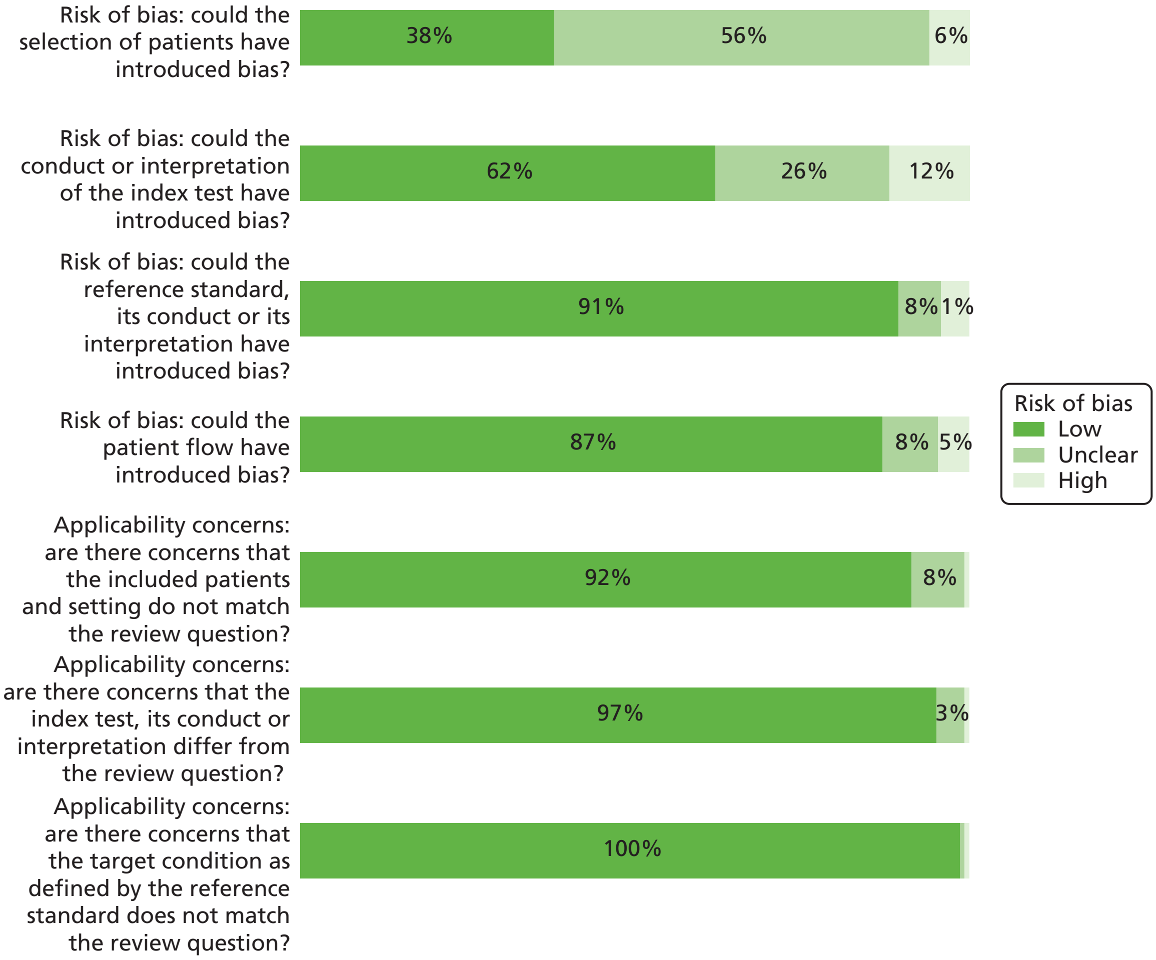


FIGURE 20 Summary of risk of bias and applicability assessment.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 16 The QUADAS-2 risk-of-bias decisions

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Abdel Razek (2017)[37]	Unclear. Not stated if enrolment was consecutive	Low	Low	Low
Al-Khawari (2005)[38]	Unclear. Unclear if participants were included consecutively and limited reporting of selection criteria	Low	Low	Low
Al-Sheikh (1985)[39]	Low	High. Index test results interpreted with knowledge of the results of the reference standard	Low	Low
Aragón-Sánchez (2011)[40]	Low	Low. Although lack of blinding to initial probe-to-bone test for radiography cannot be excluded	Low	Low. Although 91 patients with negative radiography and probe-to-bone test did not undergo histopathology tests (25.5%)
Aslangul (2013)[41]	Low	Low. 'There were findings compatible with OM on plain radiography in 26 of 42 (60%) of the patients who underwent bone puncture. Thus the enrolled patients were all at high risk for underlying OM.' Plain radiography was not an index test in this study	Low	Low. Although ≥ 1 year follow-up and no biopsy/culture for all 13 patients with negative index test (13/53 or 25% of total patients). All positive index test patients received microbiology
Blume (1997)[42]	Unclear. Unclear if patients were recruited consecutively and limited reporting of selection criteria	Low. Although antibiotics were used in 12/27 patients	Low	Low
Bohchelian (2002)[43]	Unclear. Unclear if consecutively enrolled and exclusion criteria NR	Low	Low. Although it is unclear what methods were used to prevent soft tissue contamination when collecting bone sample for culture	Low
Bolouri (2013)[44]	Unclear. A total of 4/42 patients had a known OM with suspected exacerbation	Unclear. Owing to uncertainty regarding blinding of staff to different index tests	Low. Although 12/35 had clinical follow-up without histology, some of which were positive	Low. 30 patients had biopsy. 12 only had clinical/imaging follow-up, some of whom were confirmed positive
Brunel (2016)[45]	Unclear. Unclear if consecutively enrolled	Low	Low	Low
Chacko (2003)[46]	Unclear. Insufficient information, retrospective study	Low	Low	Low. Although the number of patients who received only 6 ≥ months of clinical follow-up in the patients confirmed negative (23/46) is unclear

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Croll (1996)[47]	Unclear. Unclear if recruited consecutively	High. Not blinded to results of other tests	Low. Although 6 out of 27 diagnosed by clinical follow-up	Low. Although 6/27 only had clinical follow-up
Demirev (2014)[48]	Low	High. MRI and PET-CT results not blinded (in 19/27, MRI was performed before PET-CT, and after in 8/27 cases)	Low. Although 15/27 patients had histology/microbiology confirmed from bone sample (most confirmed positive). Other methods of sample collection included blood sampling (seven), pus aspiration (five), soft tissue debridement (one negative). Clinical follow-up in four patients	Low. Although 15/27 patients had histology/microbiology confirmed from bone sample (most confirmed positive). Other methods of sample collection included blood sampling (seven), pus aspiration (five), soft tissue debridement (one negative). Clinical follow-up in four patients
Enderle (1999)[49]	Low	Low	Low	Low
Erdman (1991)[50]	Low	Low	Low. A total of 29/110 patients diagnosed by clinical follow-up	Low. Although 29/56 of index test negative patients were confirmed negative by clinical follow-up only
Ertugrul (2006)[51]	Unclear. Unclear if patients were enrolled consecutively and why not all patients received both index tests (MRI: 28/31, scintigraphy: 26/31)	Unclear. No reporting on blinding to index test results	Low	Low
Ezzat (2011)[52]	High. Excluded patients with 'presence of any chronic disease with signs and symptoms similar to acute OM'	Low	Unclear. Most (23/27) patients had surgery, only two had cytology (pus aspiration) and two had 1-week clinical follow-up. It is unclear how many had histology/microbiology	Low
Familiari (2011)[53]	Unclear. Unclear if this was a consecutive or random sample, detail on exclusions lacking	Low. Several cut-off values were considered, but it does not appear that optimal cut-off values were derived a posteriori based on ROC curves	Low	Low
Filippi (2006)[54]	Low	Low	Low	Low
Filippi (2009)[55]	Low	Low	Low	Low. Although 5/17 patients only had clinical follow-up, the duration of follow-up (24 months) and results suggest a low risk of verification bias

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Franceschi (2013)[56]	Unclear. Conference abstract, insufficient information on selection	Low	Low	High. Only participants who received surgery were analysed (17 out of 39 who received PET/CT)
Gemmel (2004)[57]	High. Only patients with final diagnosis confirmed by microbiology and/or pus visible during surgery were included	Low	Low	Low
Guhlmann (1998)[58]	Unclear. Not stated if the sample was consecutive and exclusion criteria not reported	Low	Low	Low for ^{18}F-FDG-PET. High for scintigraphy, owing to exclusion of significant number of inconclusive results (14, of which nine were confirmed positive and five were confirmed negative)
Guhlmann (1998)[59]	Unclear. Unclear if consecutive and no reporting of exclusion criteria	Low	Low	Low
Hakim (2006)[60]	Unclear. Unclear if patients were enrolled consecutively and limited information on selection criteria	Unclear. Unclear if PET and SPECT results were interpreted independently and blinded	Low	Low. Only 4/34 (11.8%) did not have a biopsy
Hartmann (2007)[61]	Unclear. Unclear if consecutively enrolled	Low	Low	Low
Hazenberg (2011)[62]	Unclear (conference abstract, insufficient detail). Unclear because of question 1	Unclear (conference abstract, insufficient detail)	Low	Low (based on limited data from the conference abstract)
Heiba (2017)[63]	Unclear. Not sure why the diagnosis was not included in some patients' records	Low	Low	Low. A total of 12/33 had clinical follow-up only but most were confirmed negative
Horger (2003)[64]	Unclear. Unclear if consecutive with no inappropriate exclusions	High. It appears that interpretation of scintigraphy + SPECT and SPECT/CT were not done blinded	Low	Low. Although 9/29 sites had 6 months' follow-up and no biopsy
Horger (2007)[65]	Low	High. No blinding to results of various index tests	Low. Although 15/31 had surgery/biopsy and 16/31 had clinical, microbiology and/or radiography follow-up	Low. Although 15/31 had surgery/biopsy and 16/31 had clinical, microbiology and/or radiography follow-up, and it is unclear if all confirmed positive (nine) had a biopsy

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Huang (1998)[66]	Low	Low	Low	Low. A total of 10/42 (23.8%, five confirmed positive and five confirmed negative) patients had clinical follow-up only (6 months to 3 years)
Johnson (2009)[67]	Unclear. 'Repeated examinations of the same patient were excluded from the study, as were examinations of patients who had insufficient clinical follow-up to establish the presence or absence of OM.' Unclear how many patients were excluded. Patients with repeated examinations may have been harder to diagnose	Low	Low	Low. Although negative cases confirmed by clinical follow-up (duration unknown) in 85% of cases. Positive cases confirmed by histopathology in 90% of cases
Kaim (2000)[68]	Unclear. Unclear if consecutive	Low	Low. Although 5/18 only had clinical follow-up (> 16 months)	Low. Although 5/18 only had clinical follow-up (> 16 months)
Kim (2017)[69]	Low. Although limited to those who had both scans within 1 week	Low	Low. Nine cases of surgery/histopathology, 12 cases of clinical follow-up, but all positive cases confirmed by bone biopsy	Low. Nine cases of surgery/histopathology, 12 cases of clinical follow-up, but all positive cases confirmed by bone biopsy
La Fontaine (2016)[70]	Unclear. Retrospective chart review, unclear if all eligible patients were included	Low	Low	Low
Larcos (1991)[71]	Unclear. Unclear if consecutive recruitment	Unclear. Unclear if index test results were interpreted blinded to each of the other index tests	Low. It appears that all positives and a significant proportion of negatives were confirmed by histopathology	Low. It appears that all positives and a significant proportion of negatives were confirmed by histopathology
Larson (2011)[72]	Unclear	Unclear. Unclear if index tests were interpreted blinded to one another	Low	Low. Sensitivity and specificity were reported, although the final number of confirmed positive/negative cases were not provided, therefore the 2 × 2 results could not be calculated

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Ledermann (2000)[73]	Low. Although high rate of patients with metal artefacts (11/15) and only included those with a possible relapse of infection in chronic post-traumatic OM	Unclear. Owing to lack of information on blinding to CT results	Low. Final diagnosis was established in 11 patients by intraoperative microbiological and histological examination, although potentially less accurate for clinical follow-up in 4/15 patients	Low. Most (11/15) confirmed by microbiology/ histopathology
Levine (1994)[74]	Unclear. Unclear if patients were enrolled consecutively and why there were significant discrepancies in the number of patients receiving each of the four index tests	Unclear. Owing to lack of information on blinding to reference standard and other index test results	Low. Although most confirmed negative patients did not have histopathology (11/16)	Low. Although 9/36 patients were excluded owing to 'no surgical or histological confirmation within 2 weeks of MRI or failure to resolve with nonoperative management'
Lewis (1988)[75]	Unclear. Unclear if the sample was consecutively recruited; selection criteria were not reported	Unclear. Owing to insufficient details on blinding of staff performing index tests	Low	Low. Although nine (14.8%) excluded owing to 'failure to separately examine the bony portion of a pressure sore specimen or misrouting of specimens to microbiology or gematology laboratories'
Lipman (1998)[76]	Low	Low	Low	Unclear. Unclear how many received radiography and scintigraphy
Mahendra (2017)[77]	Low	Low	Low	Low
Malcius (2009)[78]	Unclear. Unclear if consecutively selected. Exclusions or reasons for exclusions were NR	High. Not all patients had the same tests and all results were considered together: scintigraphy performed if unclear diagnosis based on early radiography and/or ultrasound, and MRI/CT performed if unclear diagnosis based on scintigraphy	Unclear. Number receiving histology/ microbiology vs. clinical follow-up unknown. Uncertainty about blinding of reference standard assessors to index test results	Unclear. Number receiving histology/ microbiology vs. clinical follow-up unknown
Mason (1989)[79]	Low. Authors suggested that there was a selection bias because only patients considered likely to need surgery were studied. Such patients are probably more likely to have infection, although they may not be easier to diagnose	Unclear. Owing to uncertainty about blinding between index tests	Low	Low

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
McCarthy (2017)[80]	High. 'Patients were excluded if they did not have a preoperative MRI, if they did not have intraoperative bone culture data, or if they underwent debridement only without reconstruction'	Low	Low	Low. Only a small proportion of participants (indeterminate results, 8.5%) were excluded from analyses
Meller (2002)[81]	Low	Low	Low. 9/11 positive locations were confirmed by histology or culture. 10/25 negative locations were confirmed negative by histology or culture. Other locations were diagnosed by MRI (*n* = 15) or clinical follow-up (*n* = 2)	Low. 9/11 positive locations were confirmed by histology or culture. 10/25 negative locations were confirmed negative by histology or culture. Other locations were diagnosed by MRI (*n* = 15) or clinical follow-up (*n* = 2)
Miki (2015)[82]	Unclear. Retrospective study, only patients with surgery. Unclear if all eligible patients were recruited	Unclear. No information on blinding to reference standard results	Low	Low
Modic (1985)[83]	Unclear. Unclear if consecutively recruited and limited reporting of selection criteria	High. Owing to lack of blinding to other index tests	Low. Although 14/37 appeared to have not undergone biopsy. Clinical follow-up ranged from 6 weeks to 15 months and included antibiotics use	Low. Although 14/37 appeared to have not undergone biopsy. Clinical follow-up ranged from 6 weeks to 15 months and included antibiotics use
Morales Lozano (2010)[84]	Unclear. Unclear if selection was consecutive	Unclear. No evidence that interpretation of results of index tests were blinded to each other	Low	Low
Morrison (1993)[85]	Unclear. Unclear if consecutively recruited, limited reporting of selection criteria	Unclear. Owing to lack of information on blinding between index test assessments	Low	Low. Although small risk of verification bias (16.3% of patients did not receive histopathology), and it was not clear when histopathology was performed
Morrison (1998)[86]	Unclear. Unclear if significant number of participants were excluded because of lack of participants with no confirmed diagnosis	Low	Low. 34 were confirmed positive by biopsy, although nine confirmed positive with clinical follow-up including blood cultures, and 30 confirmed negative by clinical follow-up (rapid improvement after conservative management)	Low. 34 were confirmed positive by biopsy, although nine confirmed positive with clinical follow-up including blood cultures, and 30 confirmed negative by clinical follow-up (rapid improvement after conservative management)

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Nath (1992)[87]	Unclear. Unclear whether or not the patients were recruited consecutively and because of limited reporting of selection criteria	Unclear. Unclear blinding to reference standard and between index tests	Low. Although unclear if all samples were taken from fluid from the bone itself	Low. Only three had clinical follow-up only
Nawaz (2010)[88]	Low	Unclear. Unclear if staff were blinded to the results of each of the three index tests performed	Low	Low. Although 33.6% of all patients received histopathology. Most patients with final negative diagnosis (94.5%) were confirmed by clinical examination only. Most (85.2%) with final positive diagnoses were confirmed by histopathology
Newman (1991)[89]	Low	Unclear. Radiography and scintigraphy were interpreted in a blinded fashion, but it is unclear if interpretation of each of the two scintigraphy methods was performed blinded. Bone scans were performed after WBC scans	Low	Low. Although between 35 and 39 out of 41 eligible ulcers were included in the analyses across the different index tests. Reasons for exclusion were not reported
Newman (1992)[90]	Unclear. Unclear why only 12 of 23 patients were included. The scintigraphy data are a subset of Newman 1991[89]	Low	Low	Low
Nigro (1992)[91]	Low. Unclear if selection was consecutive, although it appears that all eligible patients were included	High. Readers were not blinded to the results of the other index tests	Unclear. Duration of clinical follow-up for 13/47 examinations (of which 11 were confirmed negative) was not reported	Low. Although 13/47 had only clinical follow-up of unknown duration, including around 50% (11/21 for MRI) of confirmed negative cases
Park (1982)[92]	High. Only diabetic patients who had radiographs, three-phase bone scans, and histological confirmation of the diagnosis, and surgery < 4 weeks after scan	Low	Low	Low. Although three patients with absent flow were excluded from the analysis for scintigraphy
Rastogi (2016)[93]	Low	Unclear. It is unclear if the results of each of the two index tests were interpreted blinded and independently	Low. Although some concerns about the accuracy of Jamshidi needle biopsy	Low. Reasons for exclusion appeared appropriate

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Remedios (1998)[94]	Unclear. Unclear if consecutively recruited	Low. The comparison of the two scintigraphy tests may have biased the scintigraphy assessment, although the actual comparison between MRI and scintigraphy was blinded	Low. All four positive cases were confirmed by biopsy, although four patients were confirmed negative by clinical follow-up only, and one confirmed negative by biopsy	Low. All four positive cases were confirmed by biopsy, although four patients were confirmed negative by clinical follow-up only, and one confirmed negative by biopsy
Rozzanigo (2009)[95]	Unclear. Retrospective study, unlikely to be consecutive. Limited reporting of selection criteria	High. Radiography performed before MRI to inform conduct of MRI (although the study was primarily designed to evaluate MRI, not to compare radiography with MRI)	Unclear. Diagnostic method and duration of follow-up unknown for confirmation of 5/16	Unclear. Diagnostic method and duration of follow-up unknown for confirmation of 5/16
Şanlı (2011)[96]	Low	Low	Unclear. Unclear how many received histopathology or microbiology: histopathology, or microbiology or observation at surgery, or ≥ 1 year follow-up. Unclear if reference standard interpretation was blinded to index test results	Unclear. Unclear how many received histopathology or microbiology; histopathology, or microbiology or observation at surgery, or ≥ 1-year follow-up
Sarikaya (2003)[97]	Unclear. Unclear if consecutively recruited	Low. For radiography: interpreted blinded to scintigraphy results. High. Owing to lack of blinding between MDP and scintigraphy	Low. Culture with/without histology conducted in 36/55 lesions. Histology and culture conducted in all positive cases. Clinical follow-up only (minimum 4 months) conducted in 19/55	Low. Although clinical follow-up only (minimum 4 months) conducted in 19/55 lesions. 26/55 received culture with/without histology. All positive cases were confirmed by histology
Schlung (2016)[98]	Unclear. Retrospective selection of subgroup of patients with MRI findings	Low	Low	Low
Schwegler (2008)[99]	Low	Low	Low. Although all 13/20 negative cases were confirmed by clinical 2-year follow-up only	Low. Although all 13/20 negative cases were confirmed by clinical 2-year follow-up only
Seabold (1990)[100]	Unclear. Unclear if consecutively recruited patients and limited details about selection criteria	Low. Results of scintigraphy and MRI blinded, although MRI was interpreted in conjunction with radiography	Low	Unclear. Only 7/14 participants were included for the analysis of MRI accuracy, with reasons NR

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Seabold (1995)[101]	Unclear. Retrospective study, insufficient detail on selection	Low	Low	High. Equivocal test results were excluded from analyses (two for SPECT and six for CT in the initial test phase) and MRI was only conducted and analysed in 11 out of 25 patients
Segall (1989)[102]	High. Only those who had histology result within 3 weeks of scan: 83 patients had three-phase bone scan, of whom 23 met the final inclusion criteria. Risk of systematic differences between included and excluded participants	Low	Low	Low
Shemesh (2015)[103]	Unclear. Retrospective study, limited reporting of selection process	Low	Low	Low
Unger (1988)[104]	Low	Low	Unclear. 14/35 (40%) only had clinical follow-up, and it is not clear if any of the patients with surgery had histology/microbiology	Unclear. 14/35 (40%) only had clinical follow-up, and it is not clear if any of the patients with surgery had histology/microbiology
van Vliet (2018)[105]	Low, although retrospective	Unclear. Unclear if multiple diagnostic cut-off points were prespecified	Low. Although 1-year follow-up only was used in some (7/30, including three confirmed positive)	Low
Weber (1995)[106]	Unclear. Retrospective and limited reporting of selection criteria	Low	Low	Low. Despite some limitations (exclusion of three patients, two patients with no culture/histology)
Weinstein (1993)[107]	Low	Low	Low	High. 15 patients with no histology and clinical follow-up only were excluded from the analyses
Wenter (2016)[108]	Unclear	Low for qualitative results High for quantitative test results, owing to post hoc choice of optimal cut-off point based on ROC analysis from the study population	Low	Low
Weon (2000)[109]	Low	Low	Low. Although duration of clinical follow-up was not reported in patients who did not have histology/microbiology (11/37)	Low. Although most confirmed negative patients (10/13) only received clinical follow-up. Duration and criteria for clinical follow-up were not reported

TABLE 16 The QUADAS-2 risk-of-bias decisions (*continued*)

First author and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Williamson (1989)[110]	Unclear. Only small number of patients with recent radiography and scintigraphy	Unclear. No evidence of blinding between index tests and with reference standard. Plausible that the CT results were not interpreted blinded to radiography and scintigraphy	Low. Although only clinical follow-up was used for patients with negative CT (3/7)	Low. Although only clinical follow-up was used for patients with negative CT (3/7)
Yang (2016)[111]	Low	Low	High. Unlikely that clinical follow-up (reference standard for 22/48 patients) was done blinded to PET test results. Concerns about the reliability of follow-up as the reference standard: patients were considered OM positive if resolution of symptoms was observed after systematic antibiotic treatment	High. Clinical follow-up was the reference standard for 22/48 (45.8%). Patients were considered OM positive if resolution of symptoms was observed after systematic antibiotic treatment
Yuh (1989)[112]	Low	Low	Low. Nearly all positive cases were confirmed by histopathology. 14/24 patients had histopathology. However, 10 had clinical follow-up only (unknown duration), and it is unclear if reference standard (including follow-up) was performed blinded to index test results	Low. Nearly all positive cases were confirmed by histopathology. 14/24 patients had histopathology. However, 10 had clinical follow-up only (unknown duration)
Zaiton (2014)[113]	Low	Low	Low	Low

NR, not reported; OM, osteomyelitis.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 17 The QUADAS-2 applicability decisions

First author and year of publication	Are there concerns that the included patients and setting do not match the review question?	Are there concerns that the index test, its conduct or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?
Abdel Razek (2017)[37]	Low	Low	Low
Al-Khawari (2005)[38]	Low	Low	Low
Al-Sheikh (1985)[39]	Low	Low	Low
Aragón-Sánchez (2011)[40]	Low. Note: diabetic foot referral centre, high prevalence of OM	Low	Low
Aslangul (2013)[41]	Low	Low	Low
Blume (1997)[42]	Low	Low	Low
Bohchelian (2002)[43]	Low	Low	Low
Bolouri (2013)[44]	Low	Low	Low
Brunel (2016)[45]	Low. Although extrapolation to patients with no spinal chord injury may be limited	Low	Low
Chacko (2003)[46]	Low	Low	Low
Croll (1996)[47]	Low	Low	Low
Demirev (2014)[48]	Low	Low	Low
Enderle (1999)[49]	Low	Low	Low
Erdman (1991)[50]	Low	Low	Low
Ertugrul (2006)[51]	Low	Low	Low
Ezzat (2011)[52]	Low. Although excluded patients with 'presence of any chronic disease with signs and symptoms similar to acute OM'	Low	Low
Familiari (2011)[53]	Low	Low	Low
Filippi (2006)[54]	Low	Low	Low
Filippi (2009)[55]	Low	Low	Low
Franceschi (2013)[56]	Unclear. Conference abstract, insufficient information about participants	Low	Low
Gemmel (2004)[57]	Unclear	Low	Low
Guhlmann (1998)[58]	Low	Low	Low
Guhlmann (1998)[59]	Low	Low	Low
Hakim (2006)[60]	Low	Low	Low
Hartmann (2007)[61]	Low	Low	Low
Hazenberg (2011)[62]	Low	Low	Low
Heiba(2017)[63]	Low	Low	Low
Horger (2003)[64]	Low	Low	Low
Horger (2007)[65]	Low	Low	Low
Huang (1998)[66]	Low	Low	Low
Johnson (2009)[67]	Low	Low	Low

TABLE 17 The QUADAS-2 applicability decisions (*continued*)

First author and year of publication	Are there concerns that the included patients and setting do not match the review question?	Are there concerns that the index test, its conduct or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?
Kaim (2000)[68]	Low. Although highly selected patient population: only long-standing, post-traumatic OM with suspected recurrent OM	Low	Low
Kim (2017)[69]	Low	Low	Low
La Fontaine (2016)[70]	Low	Low	Low
Larcos (1991)[71]	Low	Low	Low
Larson (2011)[72]	Unclear. Only some may have been referred for suspected OM (numbers unknown)	Low	Low
Ledermann (2000)[73]	Low. No significant concerns, although high proportion of patients with metal artefacts (11/15) and highly selected patient group. Only included those with a possible relapse of infection in chronic post-traumatic OM	Unclear. MRIs performed in a high proportion of patients with metal artefacts (11/15). Of these 11, four rated as 'metal artefact renders film interpretation considerably more difficult; two rated as 'metal artefact impairs complete evaluation of film because of major obscuring artefacts'	Low
Levine (1994)[74]	Low. Although limited reporting of patient characteristics	Low	Low
Lewis (1988)[75]	Unclear. Selection criteria and participant characteristics NR	Low	Low
Lipman (1998)[76]	Low	Low	Low
Mahendra (2017)[77]	Low	Unclear. No reporting on use of antimicrobials before MRI	Low
Malcius (2009)[78]	Low	Low	Low
Mason (1989)[79]	Low	Low	Low
McCarthy (2017)[80]	Low	Low	Low
Meller (2002)[81]	Low	Low	Low
Miki (2015)[82]	Low	Low	Low
Modic (1985)[83]	Unclear	Low	Low
Morales Lozano (2010)[84]	Low. Although high prevalence of OM (79.5%)	Low	Low
Morrison (1993)[85]	Low. All suspected OM	Low	Low
Morrison (1998)[86]	Low	Low	Low
Nath (1992)[87]	Low	Low	Low
Nawaz (2010)[88]	Low	Low	Low
Newman (1991)[89]	Low	Low	Low
Newman (1992)[90]	Low	Low	Low

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 17 The QUADAS-2 applicability decisions (*continued*)

First author and year of publication	Are there concerns that the included patients and setting do not match the review question?	Are there concerns that the index test, its conduct or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?
Nigro (1992)[91]	Low	Low	Low
Park (1982)[92]	Low	Low	Low
Rastogi (2016)[93]	Low	Low	Low
Remedios (1998)[94]	Low	Low	Low
Rozzanigo (2009)[95]	Low	Low	Low
Şanlı (2011)[96]	Low	Low	Low
Sarikaya (2003)[97]	Low. Patients with initial normal bone scan were excluded	Low	Low
Schlung (2016)[98]	Low	Low	Low
Schwegler (2008)[99]	Low	Low	Low
Seabold (1990)[100]	Low	Low	Low
Seabold (1995)[101]	Unclear. Retrospective study, insufficient detail on selection	Low	Low
Segall (1989)[102]	Low	Low	Low
Shemesh (2015)[103]	Low	Low	Low
Unger (1988)[104]	Low	Low	Low
van Vliet (2018)[105]	Low. Although only for differentiation between aseptic and septic delayed union in lower extremity	Low	Low
Weber (1995)[106]	Low. Although combines preoperative and follow-up scans	Low	Low
Weinstein (1993)[107]	Low	Low	Low
Wenter (2016)[108]	Low	Low	Low
Weon (2000)[109]	Low	Low	Low
Williamson (1989)[110]	Low	Low	Low
Yang (2016)[111]	Low	Low	Low
Yuh (1989)[112]	Low	Low	Low
Zaiton (2014)[113]	Low. Plain radiography (no radiological changes suggesting OM), patients scheduled for surgical management	Low	Low

NR, not reported; OM, osteomyelitis.

Appendix 3 Participant selection criteria in diagnostic accuracy studies

TABLE 18 Participant selection criteria in diagnostic accuracy studies

First author and year of publication	Criteria	
	Inclusion	Exclusion
Abdel Razek (2017)[37]	Suspected acute diabetic foot OM	Previous foot surgical intervention
Al-Khawari (2005)[38]	Suspected ankle and/or foot infection	NR
Al-Sheikh (1985)[39]	Suspected bone infection	NR
Aragón-Sánchez (2011)[40]	Diabetic foot infection	Toe necrosis caused by critical ischaemia
Aslangul (2013)[41]	Chronic foot ulcer with over bony prominence, relapsing infection, or slow healing despite adequate blood flow and care	Received systemic antibiotic in past 14 days or requiring surgery
Blume (1997)[42]	Pre-existing pedal abnormalities and clinical suspicion of OM	NR
Bohchelian (2002)[43]	Diabetic foot ulcer	NR
Bolouri (2013)[44]	High suspicion of OM of the jaw	NR
Brunel (2016)[45]	Pressure ulcer (stage III/IV) with unfavourable evolution despite optimal management, referred for surgical debridement	Antibiotics within 2 weeks before biopsies, or no biopsy
Chacko (2003)[46]	Soft tissue infection overlying bone, including decubitus ulcers, cellulitis or foot ulcer (87.5%), and recent surgery (41.1%)	NR
Croll (1996)[47]	Admitted with diabetic foot infection	Obvious gangrene or a fetid foot requiring immediate surgery
Demirev (2014)[48]	All consecutive patients who underwent ^{18}F-FDG-PET/CT as well as MRI scanning for diagnosis of OM. Referred because of inconclusive previous evaluations test (including radiography), or inconclusive standard radiography and CT, or to identify location and expansion of infection before surgery	NR
Enderle (1999)[49]	Patients with type 2 diabetes only, suspected of chronic OM with a foot lesion $\geq$ grade 2 in accordance with the classification of Wagner (Wagner[173]). Impaired wound healing, despite pathogenesis-adapted therapy for $>$ 4 weeks	Patients with acute infection
Erdman (1991)[50]	Suspected acute or chronic OM, with clinical follow-up or histopathologic data	Motion artefact or metallic artefact
Ertugrul (2006)[51]	Suspected diabetic foot lesion with Wagner grade of $\geq$ 3 ulcer	NR
Ezzat (2011)[52]	Children with clinical suspicion of AOM, negative or equivocal conventional plain radiography and presenting within 2 weeks of symptoms onset	COM, 'presence of any chronic disease with signs and symptoms similar to acute OM'

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 18 Participant selection criteria in diagnostic accuracy studies (*continued*)

First author and year of publication	Criteria	
	Inclusion	Exclusion
Familiari (2011)[53]	High clinical suspicion of foot OM based on the presence of signs and symptoms of infection (exposed bone, $n = 7$)	NR
Filippi (2006)[54]	Suspected OM	NR
Filippi (2009)[55]	Diabetes, foot lesions, abnormal laboratory tests and clinical signs of infection	NR
Franceschi (2013)[56]	Chronic diabetic foot infection	NR
Gemmel (2004)[57]	Suspected spinal infection after previous spinal surgery with a final diagnosis based on microbiology and/or intraoperative findings of macroscopic pus	Recently undergone surgery (no patients with interval of < 2 months were included)
Guhlmann (1998)[58]	Recurrent OM or symptoms of OM for > 6 weeks. Trauma or surgery dating back at least 2 years	NR
Guhlmann (1998)[59]	Suspected COM with symptoms of infection lasting > 6 weeks or presence of recurrent OM	NR
Hakim (2006)[60]	Provisional chronic suppurative OM diagnosis referred for scintigraphy, with no previous COM diagnosis	NR
Hartmann (2007)[61]	Suspected COM with symptoms of infection lasting > 6 weeks or presence of recurrent OM, with histopathology or microbiology results	Trauma or surgical intervention on affected bone during the 6 months prior to the PET/CT examination
Hazenberg (2011)[62]	Persistent OM despite antibiotics therapy	Peripheral arterial disease
Heiba (2017)[63]	Clinically suspected OM	No SPEC/CT data or definitive diagnosis
Horger (2003)[64]	Suspected reactivated post-traumatic COM	NR
Horger (2007)[65]	Clinical suspicion of acute or chronic OM	NR
Huang (1998)[66]	Paralysed with clinically suspected acute OM of the pelvis/hip, equivocal radiography	NR
Johnson (2009)[67]	MRI of the foot for evaluation of suspected OM	Repeated examinations of the same patient, insufficient clinical follow-up for OM diagnosis
Kaim (2000)[68]	Suspected relapse of bone infection in leg, recurrent OM	Patients with orthopaedic devices (although three with metal artefacts were included)
Kim (2017)[69]	Suspected OM and patients for whom time interval between the two index tests was < 1 week	NR
La Fontaine (2016)[70]	Received either MRI or SPECT-CT and a biopsy within 8 weeks for suspected diabetic foot OM	NR
Larcos (1991)[71]	Suspected pedal OM, with final diagnosis proved by surgery or clinical follow-up	Patients with multisystem disorders and inadequate documentation for the foot
Larson (2011)[72]	Stage IV pressure ulcers undergoing surgical debridement with intraoperative bone culture and prior radiography. Only some may have been referred for suspected OM	Those without surgical debridgement, or treated but lacking bone culture and/or radiography were excluded
Ledermann (2000)[73]	Suspected relapse of infection in chronic (at least 1.5 years) post-traumatic OM of lower extremities with complete MRI examination on admission	NR

TABLE 18 Participant selection criteria in diagnostic accuracy studies (*continued*)

First author and year of publication	Criteria	
	Inclusion	Exclusion
Levine (1994)[74]	Suspected OM because of diabetic foot ulcer	'No surgical or histological confirmation within 2 weeks of MRI or failure to resolve with non-operative management'
Lewis (1988)[75]	Spinal cord injury patients with pressure sores	NR
Lipman (1998)[76]	Peripheral neuropathy, high clinical suspicion of OM	NR
Mahendra (2017)[77]	Diabetic foot ulcer	'Previously operated cases with persistent ulcer'
Malcius (2009)[78]	Pain in bone area, fever, functional disorder, and/or signs of infection. Scintigraphy performed if unclear diagnosis based on early radiography and/or ultrasound, and MRI/CT performed if unclear diagnosis based on scintigraphy	NR
Mason (1989)[79]	Clinical history of suspected bone infection, soft tissue infection, or both; a completed MR imaging study; prior trauma, surgery, or both; and documented operative, pathologic, microbiologic and clinical follow-up	Incomplete MRI examination, MRI after resection, lack of follow-up data
McCarthy (2017)[80]	Surgical reconstruction of pressure ulcer with preoperative MRI and intraoperative bone cultures	No MRI, no intraoperative culture data, debridement only without reconstruction
Meller (2002)[81]	Non-diabetic, suspected COM based on positive scintigraphy	NR
Miki (2015)[82]	Diabetic foot ulcer, undergone surgery and histopathology, type 2 diabetes	NR
Modic (1985)[83]	Strong suspicion of vertebral OM based on symptoms, laboratory abnormalities, abnormal or suspicious radiographic assessment and predisposing factors	NR
Morales Lozano (2010)[84]	Diabetic patients with infected foot ulcer attending a diabetic foot clinic with suspicion of OM based on ulcer specimen culture, radiography and probe-to-bone test. Single ulcer (neuropathic or neuroischemic), previous surgery for OM or unresolved ulcer following local or antibiotic treatment	Critical ischaemia or due for an operation unrelated to OM
Morrison (1993)[85]	Suspected OM	NR
Morrison (1998)[86]	Suspected foot OM with minimum 6 weeks' follow-up	NR
Nath (1992)[87]	Pain and swelling of the affected limb, suspected OM	NR
Nawaz (2010)[88]	Diabetic foot disease and/or diabetes with suspected deep-seated infection(s) of the lower extremity(or extremities). Serum glucose levels of < 200 mg/dl	NR
Newman (1991)[89]	Diabetic foot ulcer, inpatients and outpatients	Myocardial infarction, severe peripheral vascular disease, ongoing antibiotic treatment of > 7 days or patient declined to participate

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 18 Participant selection criteria in diagnostic accuracy studies (*continued*)

First author and year of publication	Criteria	
	Inclusion	Exclusion
Newman (1992)[90]	Clinical suspicion of OM by referring physician	Myocardial infarction, severe peripheral vascular disease, ongoing antibiotic treatment for > 7 days or patient declined to participate
Nigro (1992)[91]	Patients with foot inflammation and possible OM referred for MRI	NR
Park (1982)[92]	Diabetic patients who had radiographs, three-phase bone scans, and histological confirmation of the diagnosis	Surgical procedure performed more than 4 weeks after the bone scan
Rastogi (2016)[93]	Diabetic Charcot neuroarthropathy, suspected diabetic foot OM, foot ulcer or discharging sinus	Unfit for PET/CT or MRI, previous antibiotic treatment for foot infections for > 72 hours, peripheral vascular disease
Remedios (1998)[94]	Peripheral neuropathy chronic foot ulcers and clinical signs of OM	NR
Rozzanigo (2009)[95]	Infected ulcer affecting the forefoot/midfoot/hindfoot	NR
Şanlı (2011)[96]	Suspected bone or orthopaedic implant infection based on clinical signs	NR
Sarikaya (2003)[97]	Clinically suspected OM, diabetes, non-healing ulcers or gangrene	Normal scintigraphy, antibiotics within 1 week of imaging and culture
Schlung (2016)[98]	Septic hip (presumed likely or confirmed) with joint aspirate or culture data following treatment with irrigation and debridement and antibiotics who underwent preoperative MRI	No septic hip (presumed or confirmed) with joint aspirate or culture data, no preoperative MRI
Schwegler (2008)[99]	Diabetes, non-healing chronic foot ulcer for at least 8 weeks; absence of antibiotic pretreatment and of clinical signs or symptoms of local or systemic infection	Clinical signs of local infection, antibiotic treatment at time of PET or within 1 month
Seabold (1990)[100]	High clinical suspicion of OM in or around neuropathic joint(s)	NR
Seabold (1995)[101]	Suspected cranial OM undergoing diagnostic imaging	NR
Segall (1989)[102]	Histological test for OM within 3 weeks of bone scan	NR
Shemesh (2015)[103]	Cases suspected of deep infection following osteosynthesis for fractures of the tibia, treated surgically and had preoperative PET/CT scan	NR
Unger (1988)[104]	Clinical suspicion of acute OM	NR
van Vliet (2018)[105]	Underwent ^{18}F-FDG-PET/CT at level 1 trauma centre between March 2010 to October 2014	Surgery within 3 months of PET/CT
Weber (1995)[106]	Clinically suspicious postoperative initial persistent or recurrent OM	NR
Weinstein (1993)[107]	Clinical suspicion of OM, non-healing foot ulcer or soft tissue infection of the foot	NR
Wenter (2016)[108]	Suspected chronic OM treated in major trauma centre. Pain, absence of clear clinical or laboratory markers for acute local infection, non-conclusive radiography and MRI	NR

TABLE 18 Participant selection criteria in diagnostic accuracy studies (*continued*)

First author and year of publication	Criteria	
	Inclusion	Exclusion
Weon (2000)[109]	Suspected OM	NR
Williamson (1989)[110]	Suspected OM based on radiography	NR
Yang (2016)[111]	Suspected diabetic foot OM, consecutively included	NR
Yuh (1989)[112]	Diabetes with clinical suspicion of OM and/or non-healing foot ulcers	NR
Zaiton (2014)[113]	Foot ulcers with sign of infection, normal radiography with no sign of OM, scheduled for surgical management	Recurrent or long-standing OM with evident radiographic changes, ischaemic foot ulcers, minor abrasion or laceration, presence of contraindication to surgery or MRI examination

NR, not reported; OM, osteomyelitis.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Appendix 4 Results of diagnostic meta-analyses

TABLE 19 Summary of diagnostic accuracy studies

First author and year of publication	Population	Cause	Diagnostic tests used
Al-Khawari (2005)[38]	Adults	Diabetic foot ulcers	MRI
Al-Sheikh (1985)[39]	Adults	NR	Scintigraphy and radiography
Aragón-Sánchez (2011)[40]	Adults	Diabetic foot ulcers	Radiography
Aslangul (2013)[41]	Adults	Diabetic foot ulcers	SPECT/CT
Blume (1997)[42]	Adults	Diabetic foot ulcers	Scintigraphy and radiography
Bolouri (2013)[44]	Mixed	Multiple	CT, scintigraphy, SPECT/CT and radiography
Brunel (2016)[45]	Adults	Pressure ulcers	MRI
Chacko (2003)[46]	NR	Multiple	PET
Croll (1996)[47]	Adults	Diabetic foot ulcers	MRI, scintigraphy and radiography
Demirev (2014)[48]	Adults	Multiple	MRI and PET-CT
Enderle (1999)[49]	Adults	Diabetic foot ulcers	MRI, scintigraphy, ultrasound and radiography
Erdman (1991)[50]	Mixed	Multiple	MRI
Ertugrul (2006)[51]	NR	Diabetic foot ulcers	MRI and scintigraphy
Familiari (2011)[53]	Adults	Diabetic foot ulcers	PET-CT and scintigraphy
Filippi (2006)[54]	Adults	Multiple	SPECT/CT
Filippi (2009)[55]	Adults	Diabetic foot ulcers	SPECT/CT
Franceschi (2013)[56]	Adults	Diabetic foot ulcers	PET-CT
Gemmel (2004)[57]	Mixed	Surgical	Scintigraphy and SPECT
Guhlmann (1998)[58]	Adults	Trauma or surgery	PET and scintigraphy
Guhlmann (1998)[59]	Adults	Trauma or surgery	PET
Hakim (2006)[60]	Adults	Dental infection, trauma, others	PET and SPECT
Hartmann (2007)[61]	Adults	Trauma	PET-CT
Hazenberg (2011)[62]	NR	Diabetic foot ulcers	MRI
Heiba (2017)[63]	Adults	NR	Scintigraphy and SPECT/CT
Horger (2003)[64]	Adults	Trauma	SPECT/CT
Horger (2007)[65]	Adults	Multiple	SPECT/CT ($n = 15$), MRI ($n = 6$) and radiography ($n = 20$)
Huang (1998)[66]	Adults	Skin ulceration	MRI
Johnson (2009)[67]	Adults	Diabetes, other	MRI
Kaim (2000)[68]	Adults	Trauma	MRI and scintigraphy
La Fontaine (2016)[70]	Adults	Diabetic foot ulcers	MRI and SPECT/CT
Larcos (1991)[71]	Adults	Diabetic foot ulcers	Scintigraphy and radiography
Ledermann (2000)[73]	Adults	Trauma	MRI
Levine (1994)[74]	Adults	Diabetic foot ulcers	MRI, scintigraphy and radiography

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 19 Summary of diagnostic accuracy studies (*continued*)

First author and year of publication	Population	Cause	Diagnostic tests used
Lewis (1988)[75]	Adults	Skin ulceration	CT, scintigraphy and radiography
Lipman (1998)[76]	Adults	Diabetes (85%), peripheral neuropathy (100%)	MRI
Mahendra (2017)[77]	Adults	Diabetic foot ulcers	MRI
Mason (1989)[79]	NR	Trauma (93%), other (7%)	MRI and scintigraphy
McCarthy (2017)[80]	Adults	Skin ulceration	MRI
Meller (2002)[81]	Adults	Multiple	PET and scintigraphy
Miki (2015)[82]	Adults	Diabetic foot ulcers	MRI
Modic (1985)[83]	Adults	Multiple	MRI, scintigraphy (and SPECT $n = 5$) and radiography
Morales Lozano (2010)[84]	NR	Diabetic foot ulcers	Radiography
Morrison (1993)[85]	Mixed	Multiple	MRI
Morrison (1998)[86]	Adults	Diabetes (85%), others (15%)	MRI
Nath (1992)[87]	Mixed	NR	Ultrasound and radiography
Nawaz (2010)[88]	Adults	Diabetic foot ulcers	MRI, PET and radiography
Newman (1992)[90]	NR	Diabetic foot ulcers	Scintigraphy and radiography
Newman (1991)[89]	NR	Diabetic foot ulcers	MRI and scintigraphy
Nigro (1992)[91]	Adults	Diabetic foot ulcers, foot ulcers or inflammation	MRI, scintigraphy and radiography
Park (1982)[92]	NR	Diabetic foot ulcers	Scintigraphy and radiography
Rastogi (2016)[93]	Adults	Diabetic foot ulcers	MRI and PET-CT
Remedios (1998)[94]	Adults	Diabetic foot ulcers	MRI and scintigraphy
Rozzanigo (2009)[95]	Adults	Diabetic foot ulcers	MRI and radiography
Sarikaya (2003)[97]	Adults	Diabetic foot ulcers	Scintigraphy and radiography
Schwegler (2008)[99]	Adults	Diabetic foot ulcers	MRI, PET and scintigraphy
Seabold (1990)[100]	Adults	Neuropathic osteoarthropathy	MRI and scintigraphy
Seabold (1995)[101]	Mixed	Multiple	CT, MRI and SPECT
Segall (1989)[102]	Adults	Diabetic foot ulcers	Scintigraphy and radiography
Shemesh (2015)[103]	Adults	Trauma or surgery	PET-CT
Unger (1988)[104]	Mixed	NR	MRI and scintigraphy
van Vliet (2018)[105]	Adults	Trauma and septic delayed union	PET-CT
Weber (1995)[106]	Mixed	Multiple	CT and SPECT
Weinstein (1993)[107]	Adults	Diabetic foot infections	MRI, scintigraphy and radiography
Wenter (2016)[108]	Adults	Trauma	PET-CT
Weon (2000)[109]	Mixed	Mixed	Scintigraphy and SPECT
Williamson (1989)[110]	Adults	Mixed	CT and scintigraphy
Yang (2016)[111]	Adults	Diabetic foot ulcers	PET
Yuh (1989)[112]	Adults	Diabetic foot ulcers	MRI, scintigraphy and radiography
Zaiton (2014)[113]	Adults	Diabetic foot ulcers	MRI

NR, not reported.

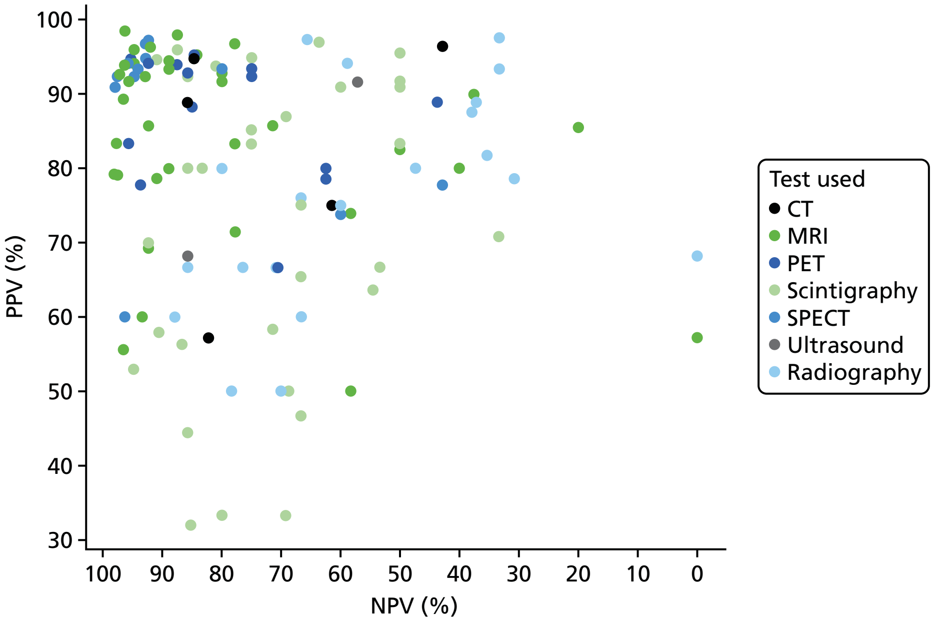


FIGURE 21 Summary of PPV and NPV in adults.

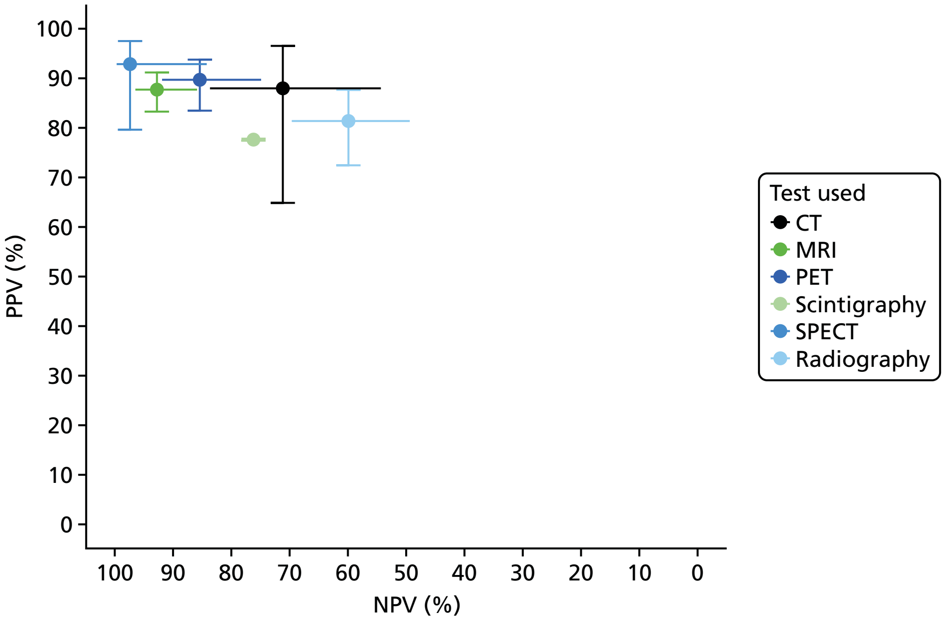


FIGURE 22 Summary of PPV and NPV from bivariate analysis.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

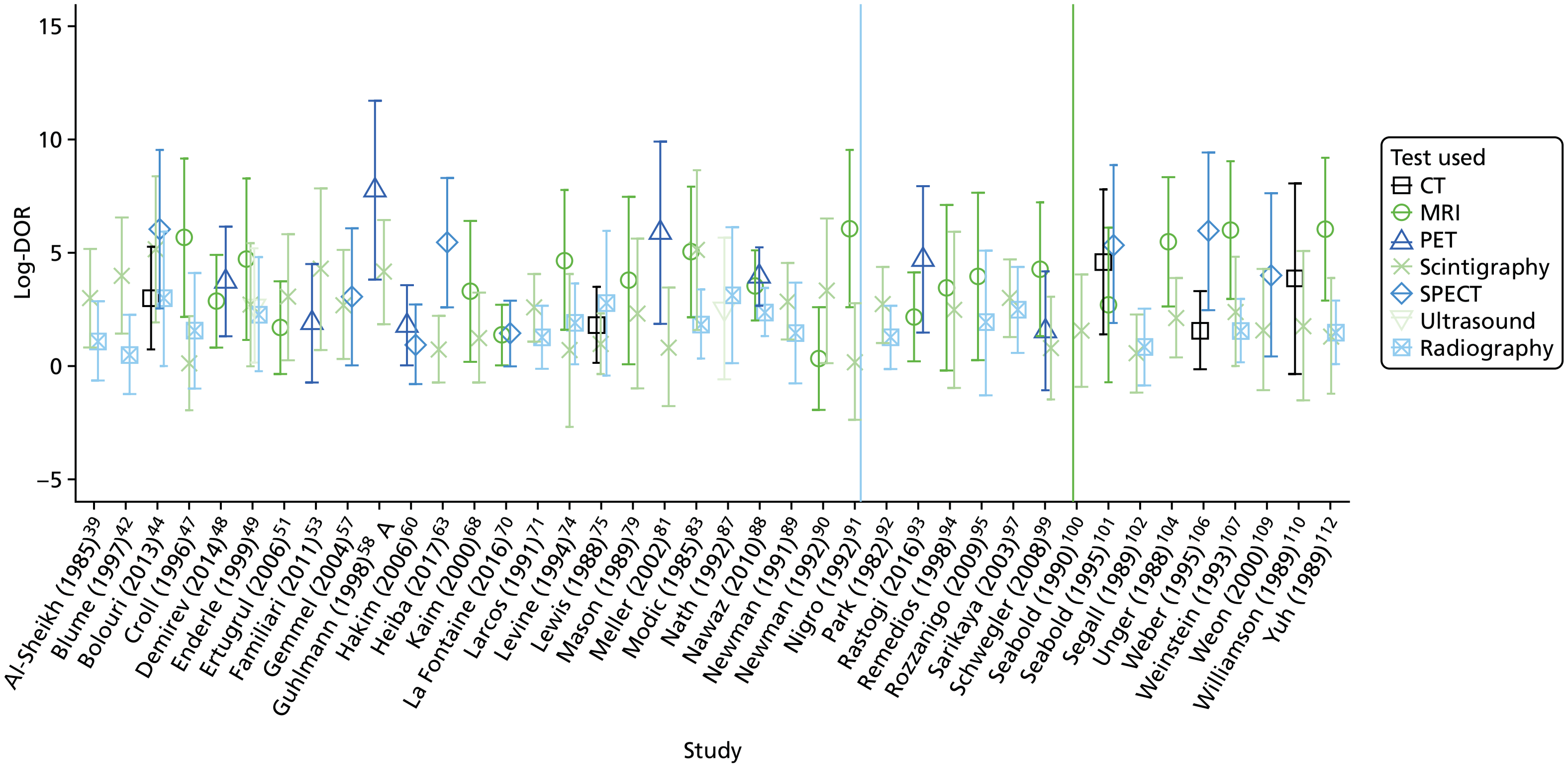


FIGURE 23 Comparisons of DORs within studies.

TABLE 20 Results from logistic regression model to compare tests

	Difference in log-odds of specificity			Difference in log-DOR		
Test	Estimate	Standard error	*p*-value	Estimate	Standard error	*p*-value
Comparison with radiography						
MRI	–0.149	0.228	0.512	2.392	0.339	0
Scintigraphy	0.83	0.198	0	0.554	0.272	0.041
CT	–0.399	0.465	0.391	1.139	0.565	0.044
SPECT	0.261	0.362	0.471	2.493	0.53	0
PET	–1.085	0.377	0.004	2.516	0.493	0
Comparison with MRI						
Scintigraphy	0.977	0.224	0	–1.838	0.336	0
Radiography	0.149	0.228	0.515	–2.2392	0.339	0
CT	–0.247	0.482	0.608	–1.258	0.609	0.039
SPECT	0.412	0.352	0.242	0.1	0.526	0.85
PET	–0.936	0.371	0.012	0.126	0.498	0.801

TABLE 21 Summary of studies of scintigraphy

First author and year of publication	Number of participants	Type or cause of osteomyelitis	Isotope and scintigraphy type
Included in main meta-analysis			
Al-Sheikh (1985)[39]	22	Multiple	^{67}Ga ^{99m}Tc-MDP ^{99m}Tc-MDP/^{67}Ga ^{111}In WBC
Blume (1997)[42]	27	Diabetic foot ulcers	^{99m}Tc-HMPAO WBC ^{99m}Tc-MDP
Bolouri (2013)[44]	42	Jaw/head	^{99m}Tc-DPD
Croll (1996)[47]	22	Diabetic foot ulcers	^{99m}Tc-MDP ^{111}In WBC
Enderle (1999)[49]	16	Diabetic foot ulcers	^{99m}Tc-MDP
Ertugrul (2006)[51]	26	Diabetic foot ulcers	^{99m}Tc-WBC and ^{99m}Tc-MDP
Familiari (2011)[53]	13	Diabetic foot ulcers	^{99m}Tc-HMPAO WBC
Gemmel (2004)[57]	22	Surgical in spine	^{99m}Tc-ciprofloxacin (Infection; Bayer plc, Newbury, UK)
Guhlmann (1998)[59]	37	Contiguous (multiple locations)	^{99m}Tc-AGAb/^{99m}Tc-MDP
Heiba (2017)[63]	33	Pelvis/hip	Not reported
Kaim (2000)[68]	19	Long bone trauma	^{99m}Tc-DPD/^{99m}Tc-Mab
Larcos (1991)[71]	49	Diabetic foot ulcers	^{99m}Tc-MDP
Larcos (1991)[71]	51	Diabetic foot ulcers	^{111}In WBC
Levine (1994)[74]	11	Not reported	^{99m}Tc-HMDP ^{111}In WBC

continued

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 21 Summary of studies of scintigraphy (*continued*)

First author and year of publication	Number of participants	Type or cause of osteomyelitis	Isotope and scintigraphy type
Lewis (1988)[75]	52	Skin ulceration	^{99m}Tc-WBC
Mason (1989)[79]	11	Long bone trauma	^{111}In WBC
Meller (2002)[81]	34	Multiple	^{111}In WBC
Modic (1985)[83]	20	Multiple	^{67}Ga ^{99m}Tc-HMDP
Newman (1991)[89]	39	Diabetic foot ulcers	^{99m}Tc-MDP ^{111}In WBC
Newman (1992)[90]	16	Diabetic foot ulcers	^{111}In WBC
Nigro (1992)[91]	39	Diabetic foot ulcers	^{99m}Tc-HMDP
Park (1982)[92]	36	Diabetic foot ulcers	^{99m}Tc-MDP
Remedios (1998)[94]	9	Diabetic foot ulcers	^{99m}Tc-NA
Sarikaya (2003)[97]	55	Diabetic foot ulcers	^{99m}Tc-MDP ^{99m}Tc-dextran
Schwegler (2008)[99]	20	Diabetic foot ulcers	^{99m}Tc-Mab
Seabold (1990)[100]	16	Neuropathic osteoarthropathy	^{111}In WBC/^{99m}Tc-MDP
Segall (1989)[102]	24	Diabetic foot ulcers	^{99m}Tc-MDP
Unger (1988)[104]	31	Multiple	^{99m}Tc-MDP
Weinstein (1993)[107]	22	Diabetic foot infections	^{99m}Tc-MDP/^{67}Ga
Weon (2000)[109]	12	Hip/knee	^{99m}Tc-HMPAO WBC
Williamson (1989)[110]	7	Foot (mixed causes)	^{99m}Tc-MDP
Yuh (1989)[112]	29	Diabetic foot ulcers	^{99m}Tc-MDP
Additional studies reporting only scintigraphy			
Glaudemans (2013)[132]	61	Multiple	^{99m}Tc-HMPAO WBC
Harwood (1999)[136]	47	Diabetic foot ulcers	^{99m}Tc-MDP ^{99m}Tc-Mab ^{99m}Tc-WBC
Kolindou (1996)[142]	107	Multiple	^{111}In WBC/^{99m}Tc-MDP
Mora (2010)[148]	94	Not reported	^{99m}Tc ciprofloxacin
Nepola (1993)[149]	44	Trauma or surgery	^{99m}Tc-MDP and ^{111}In WBC scan
Nijhof (1997)[150]	26	Multiple	^{99m}Tc-MDP ^{111}In IgG
Noriega-Alvarez (2017)[151]	27	Multiple	^{99m}Tc-HMPAO WBC
Oyen (1992)[152]	52	Multiple	^{99m}Tc-MDP and ^{111}In IgG
Sorsdahl (1993)[165]	126	Multiple	^{99m}Tc-MDP(3 phase) + ^{67}Ga

DPD, diphosphono-1,2-propanodicarboxylic acid.

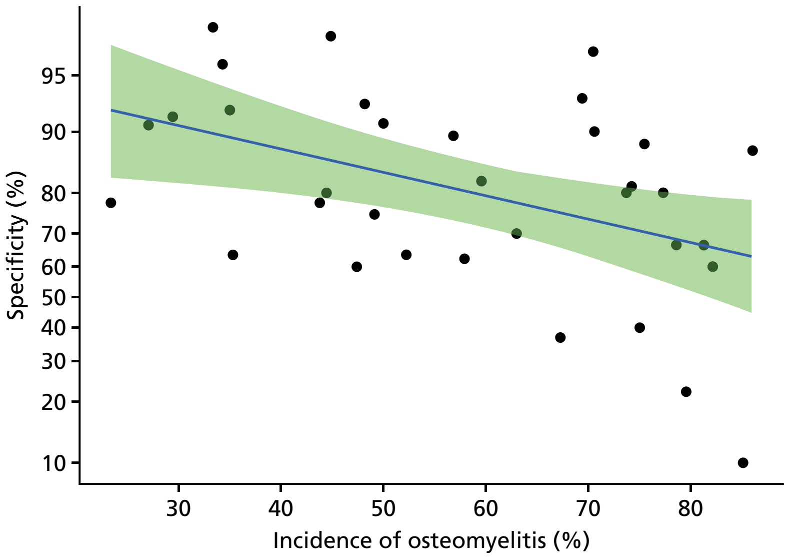


FIGURE 24 Association between specificity and incidence of osteomyelitis in MRI studies.

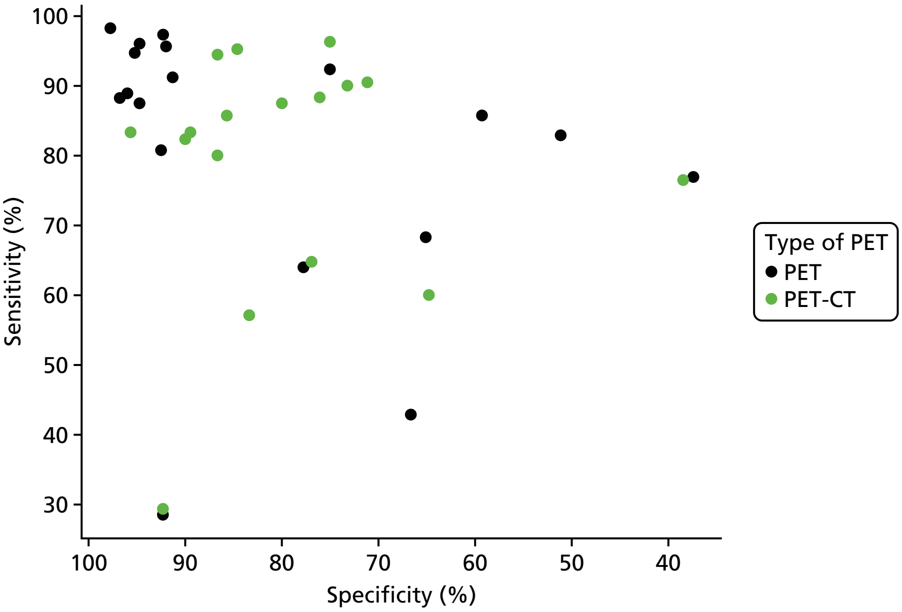


FIGURE 25 Sensitivity and specificity for PET and PET-CT studies.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 22 Results from logistic regression model to compare tests

	Difference in log-odds of specificity			Difference in log-DOR		
Test	Estimate	Standard error	*p*-value	Estimate	Standard error	*p*-value
Comparison with radiography						
MRI	–0.149	0.228	0.512	2.392	0.339	0
Scintigraphy	0.83	0.198	0	0.554	0.272	0.041
CT	–0.399	0.465	0.391	1.139	0.565	0.044
SPECT	0.261	0.362	0.471	2.493	0.53	0
PET	–1.085	0.377	0.004	2.516	0.493	0
Comparison with MRI						
Scintigraphy	0.977	0.224	0	–1.838	0.336	0
Radiography	0.149	0.228	0.515	–2.2392	0.339	0
CT	–0.247	0.482	0.608	–1.258	0.609	0.039
SPECT	0.412	0.352	0.242	0.1	0.526	0.85
PET	–0.936	0.371	0.012	0.126	0.498	0.801

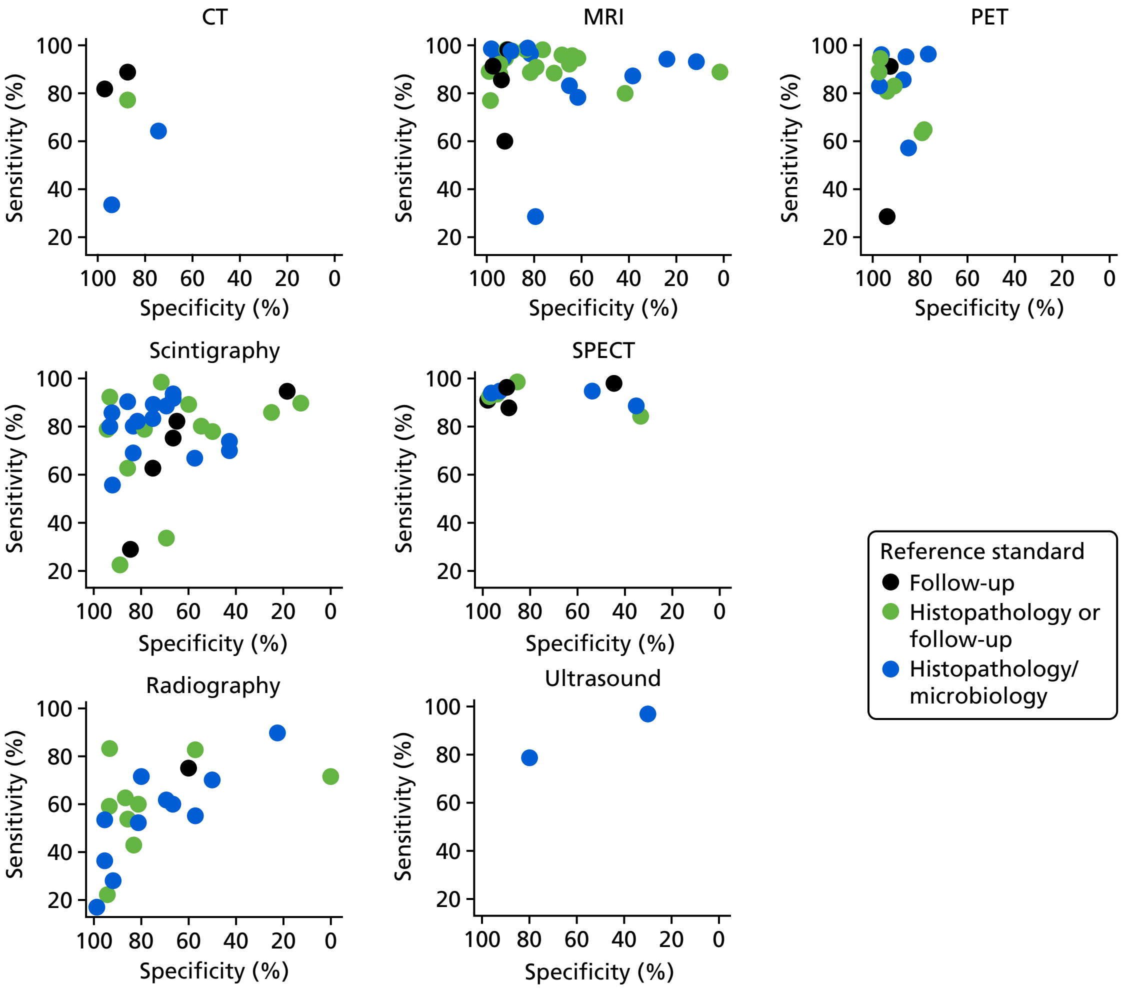


FIGURE 26 Sensitivity and specificity according to test and reference standard used.

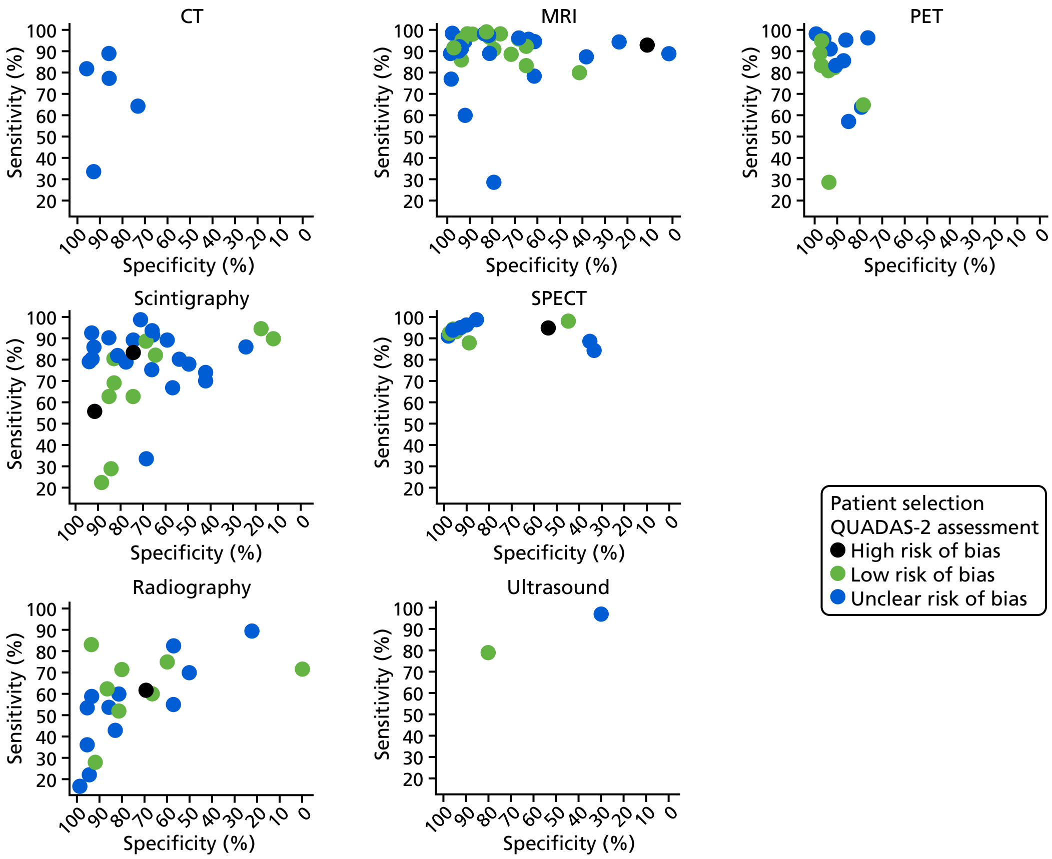


FIGURE 27 Sensitivity and specificity according to risk of bias owing to patient selection.

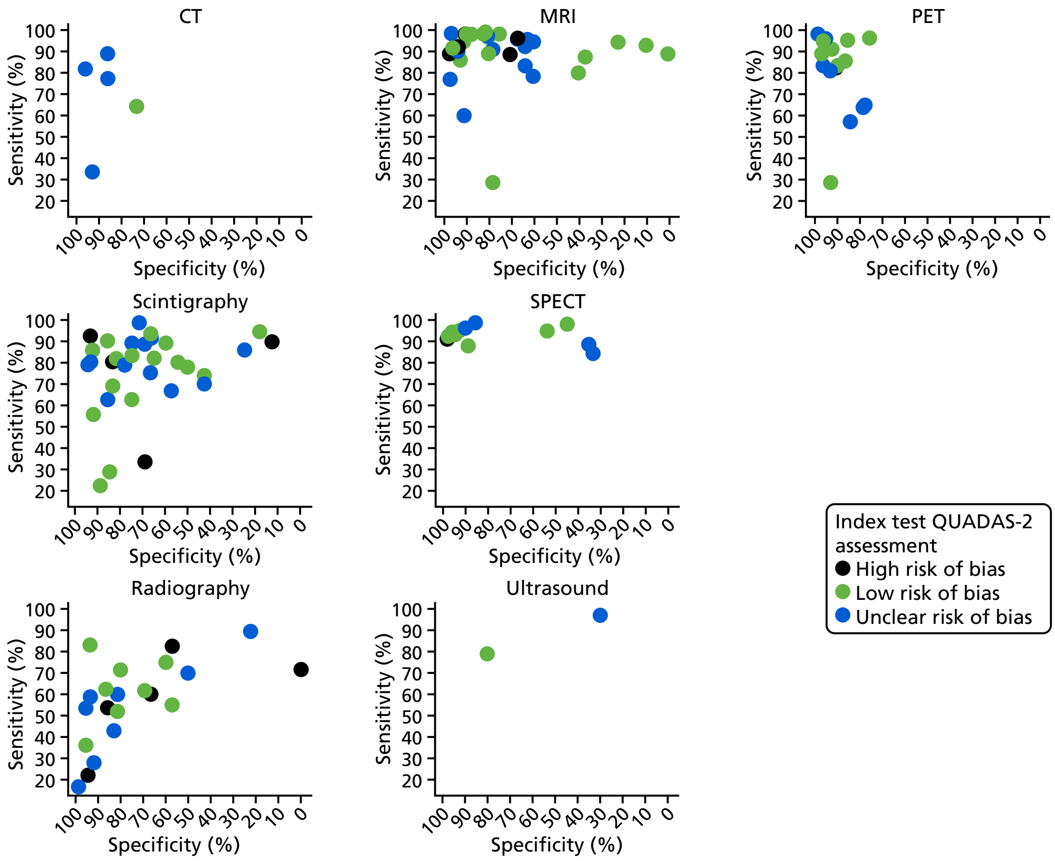


FIGURE 28 Sensitivity and specificity according to risk of bias owing to index test.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

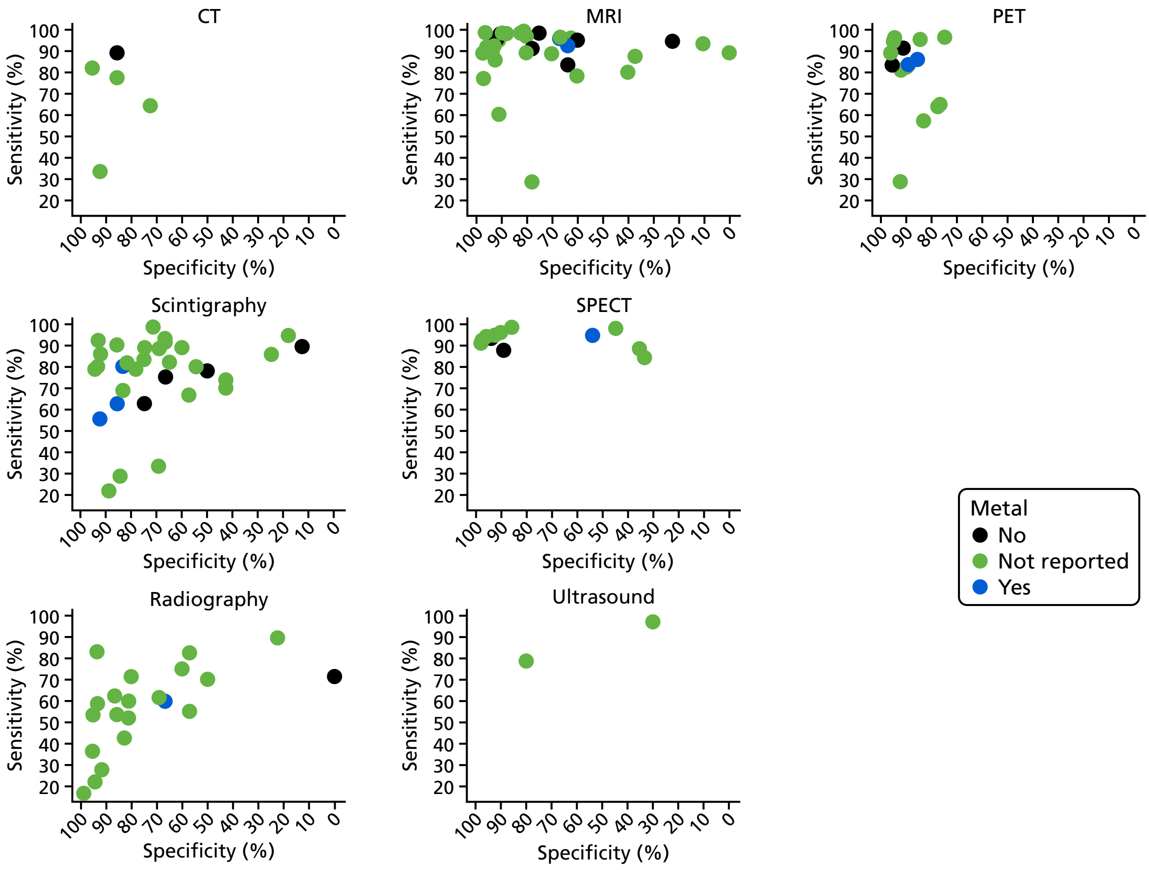


FIGURE 29 Sensitivity and specificity according to presence of indwelling metalwork.

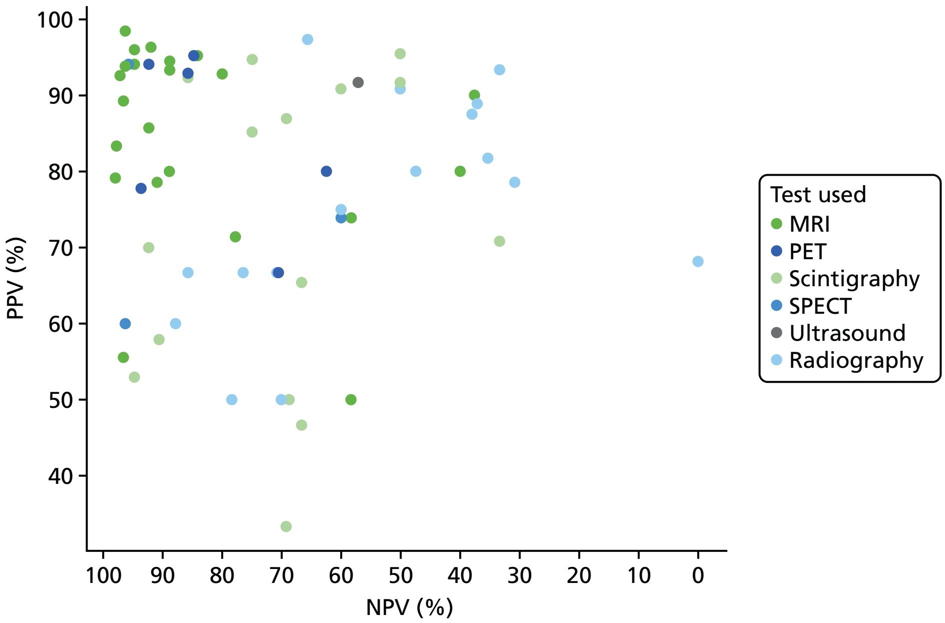


FIGURE 30 The PPV and NPV in studies of people with diabetic foot ulcers.

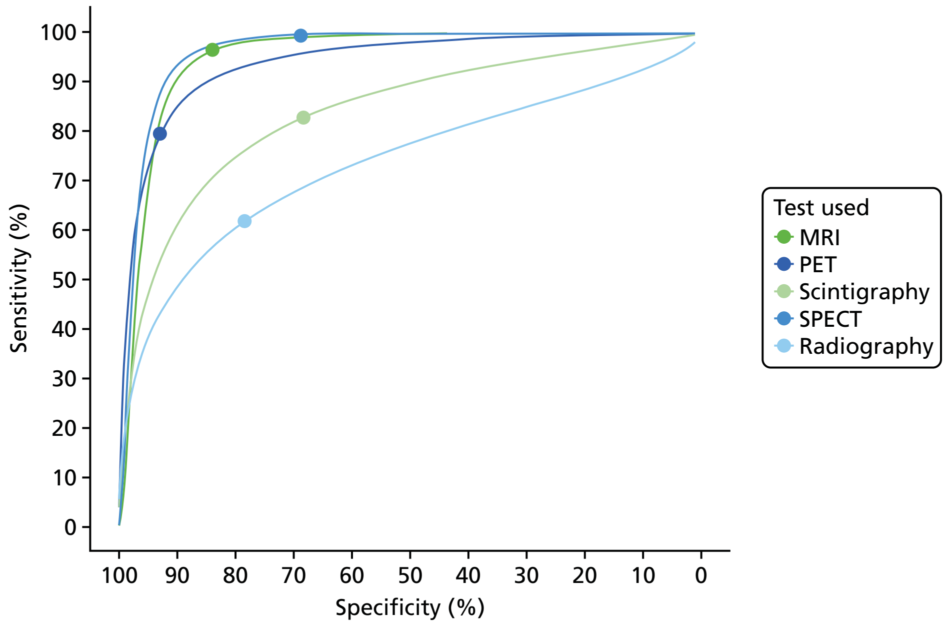


FIGURE 31 The HSROC curves in studies of people with diabetic foot ulcers.

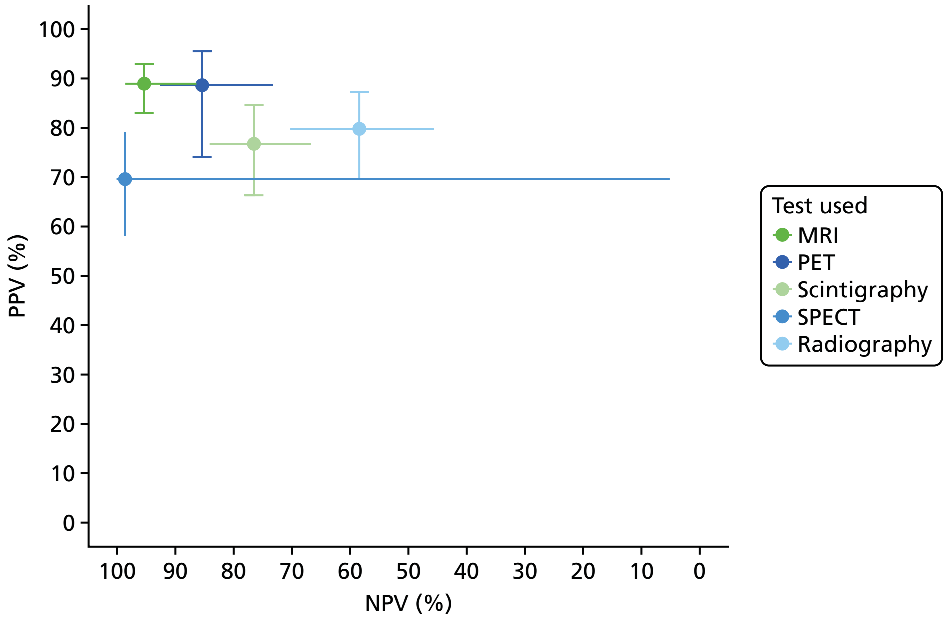


FIGURE 32 Summary PPV and NPV in studies of people with diabetic foot ulcers.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 23 Univariate meta-analyses of studies of people with diabetic foot ulcers

Test	Outcome	Estimate (%)	95% CI (%)	I^2
MRI	Sensitivity	95.76	91.83 to 97.85	0
	Specificity	81.79	69.35 to 89.91	0
	DOR	51.08	21.3 to 122.53	60
	PR	70.43	64.03 to 77.47	1
PET	Sensitivity	84.34	52.77 to 96.29	0
	Specificity	92.8	75.67 to 98.16	0
	DOR	33.91	11.75 to 97.92	12
	PR	45.88	27.81 to 75.69	36
Scintigraphy	Sensitivity	84.69	65.86 to 94.07	0
	Specificity	73.99	54.96 to 86.89	0
	DOR	8.66	4.74 to 15.83	16
	PR	73.57	64.29 to 84.2	9
SPECT	Sensitivity	95.53	75.95 to 99.31	0
	Specificity	55.09	19.26 to 86.32	36
	DOR	22.91	1.91 to 274.73	62
	PR	76.7	62.79 to 93.69	0
Ultrasound	Sensitivity	78.57	2.39 to 99.82	–
	Specificity	80	1.61 to 99.9	–
	DOR	14.67	1.16 to 185.23	–
	PR	63.16	30.87 to 129.23	–
Radiography	Sensitivity	68.91	57.55 to 78.38	11
	Specificity	77.99	63.67 to 87.76	0
	DOR	5.97	3.09 to 11.51	62
	PR	51.95	41.08 to 65.68	62

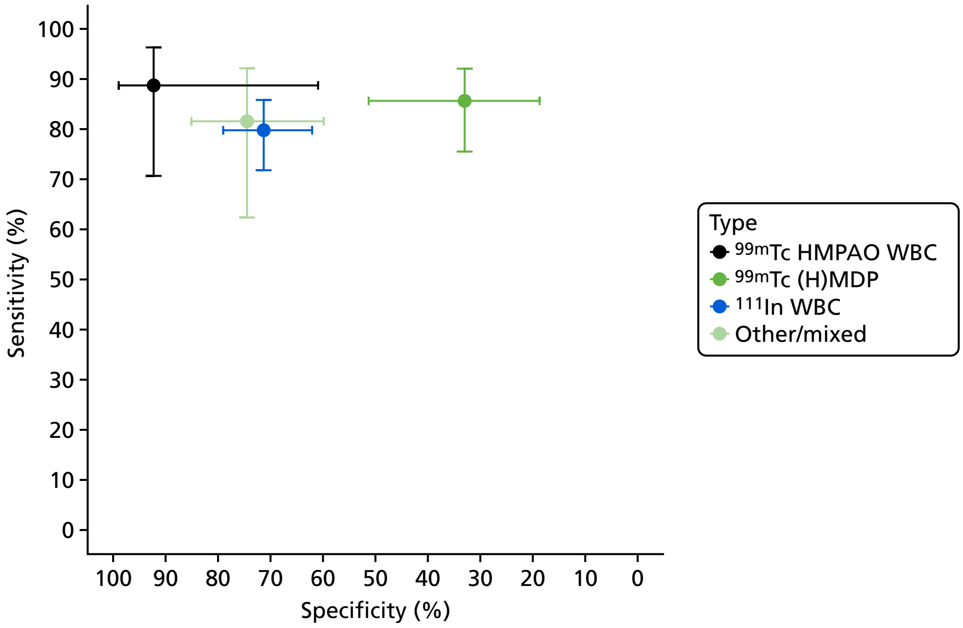


FIGURE 33 Meta-analysis of scintigraphy studies in people with diabetic foot ulcers.

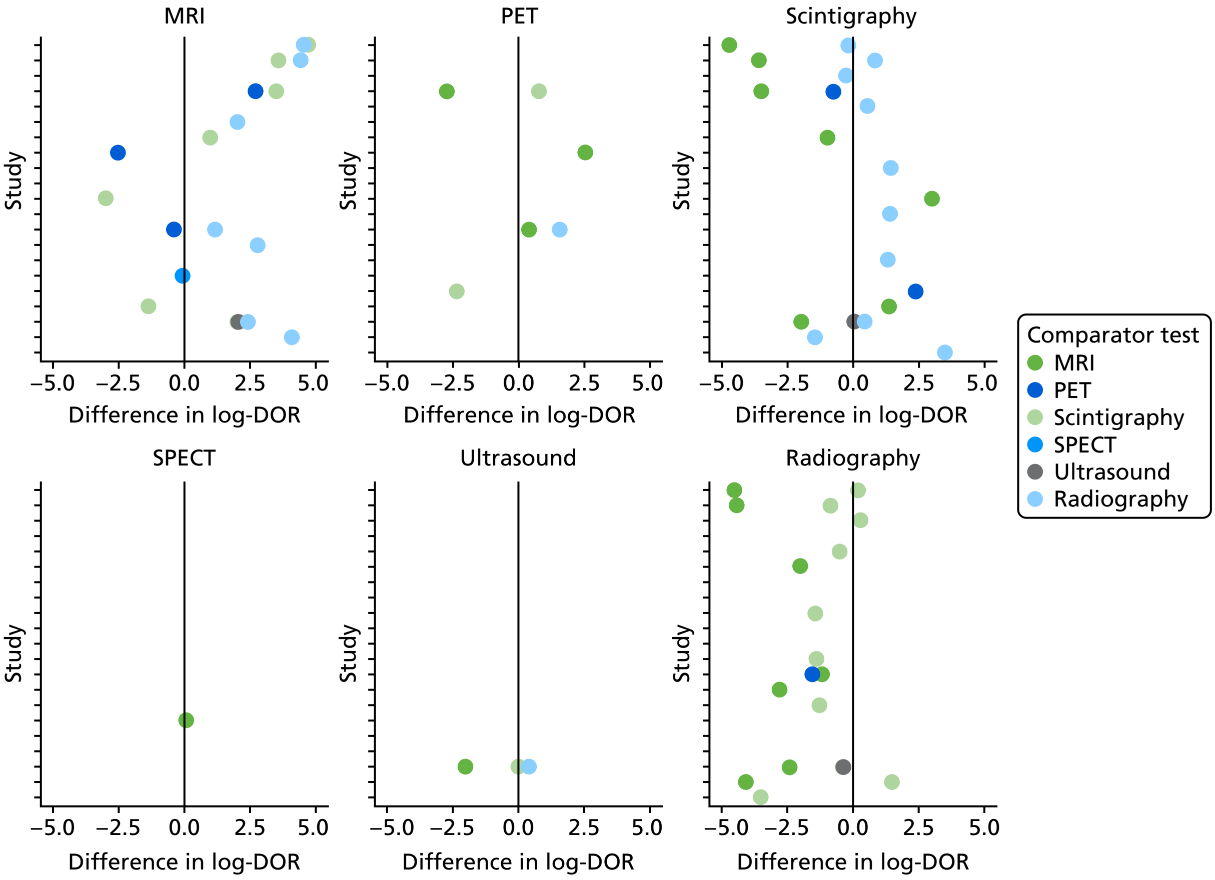


FIGURE 34 Difference in DOR between studies – patients with diabetic foot ulcers.

TABLE 24 Results of logistic regression to compare tests – diabetic foot ulcers

	Difference in log-odds of specificity			Difference in log-DOR		
Test	Estimate	Standard error	*p*-value	Estimate	Standard error	*p*-value
Comparison with radiography						
MRI	–0.175	0.255	0.493	2.188	0.366	0
Scintigraphy	0.727	0.244	0.003	0.731	0.344	0.034
SPECT	0.792	0.728	0.277	0.43	1.043	0.68
PET	–1.121	0.424	0.008	1.991	0.541	0
Comparison with MRI						
Scintigraphy	0.902	0.281	0.001	–1.458	0.389	0
Radiography	0.175	0.255	0.493	–2.188	0.366	0
SPECT	0.968	0.682	0.156	–1.758	0.992	0.076
PET	–0.946	0.429	0.028	–0.197	0.588	0.738

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

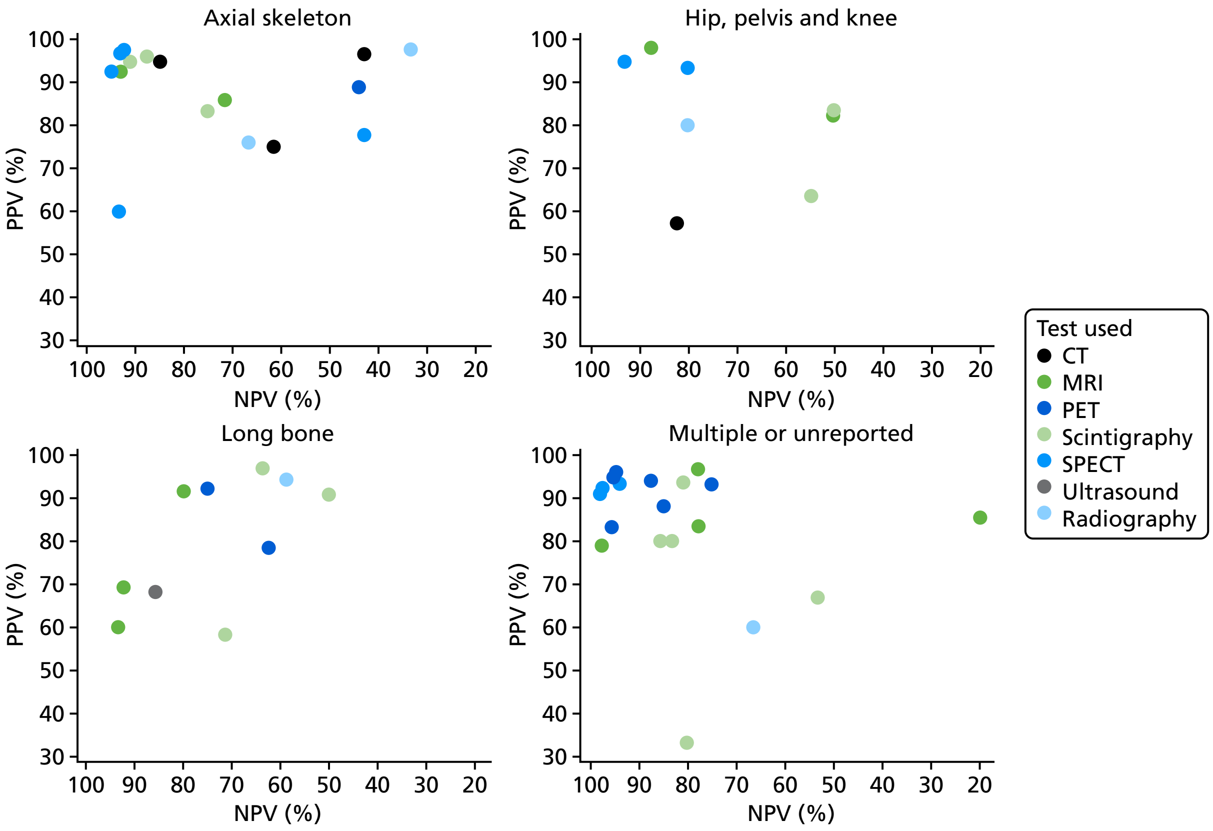


FIGURE 35 The PPV and NPV by test location.

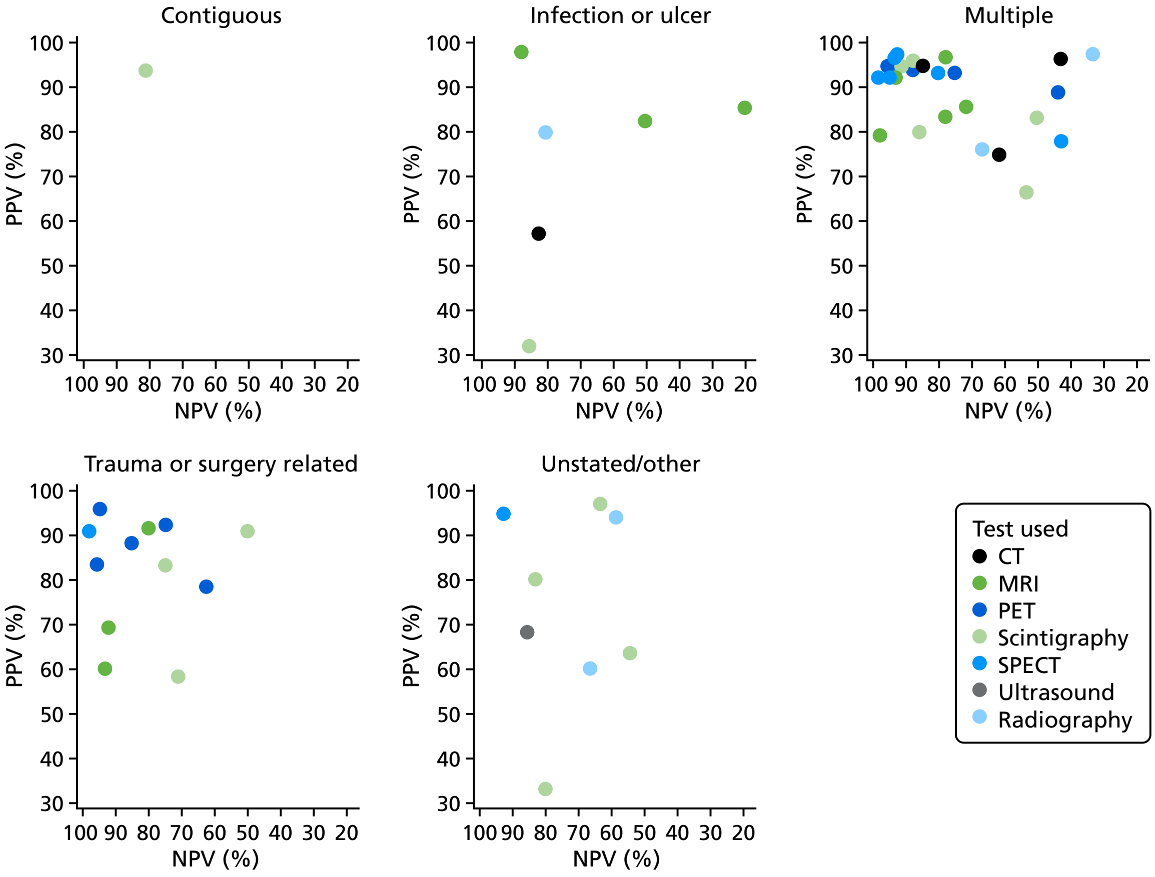


FIGURE 36 The PPV and NPV by cause of osteomyelitis.

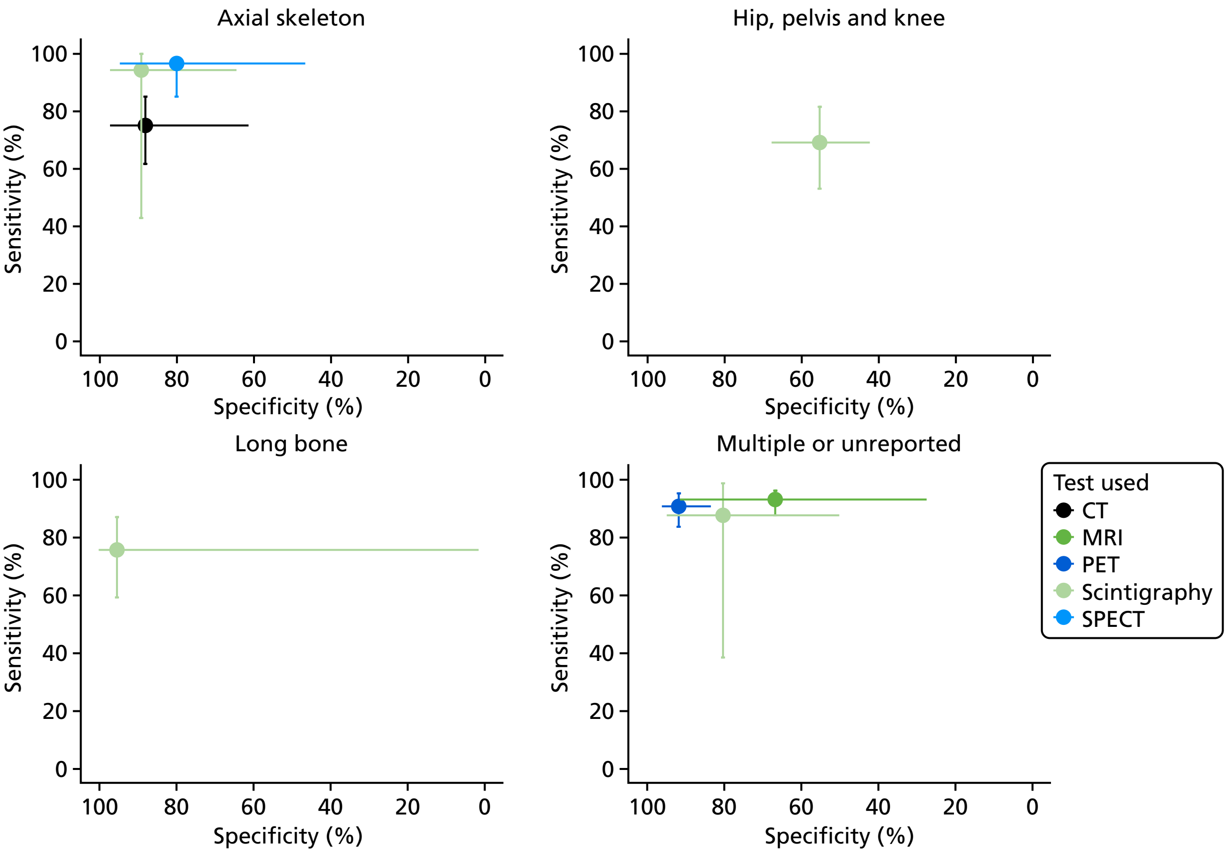


FIGURE 37 Summary sensitivity and specificity by test location.

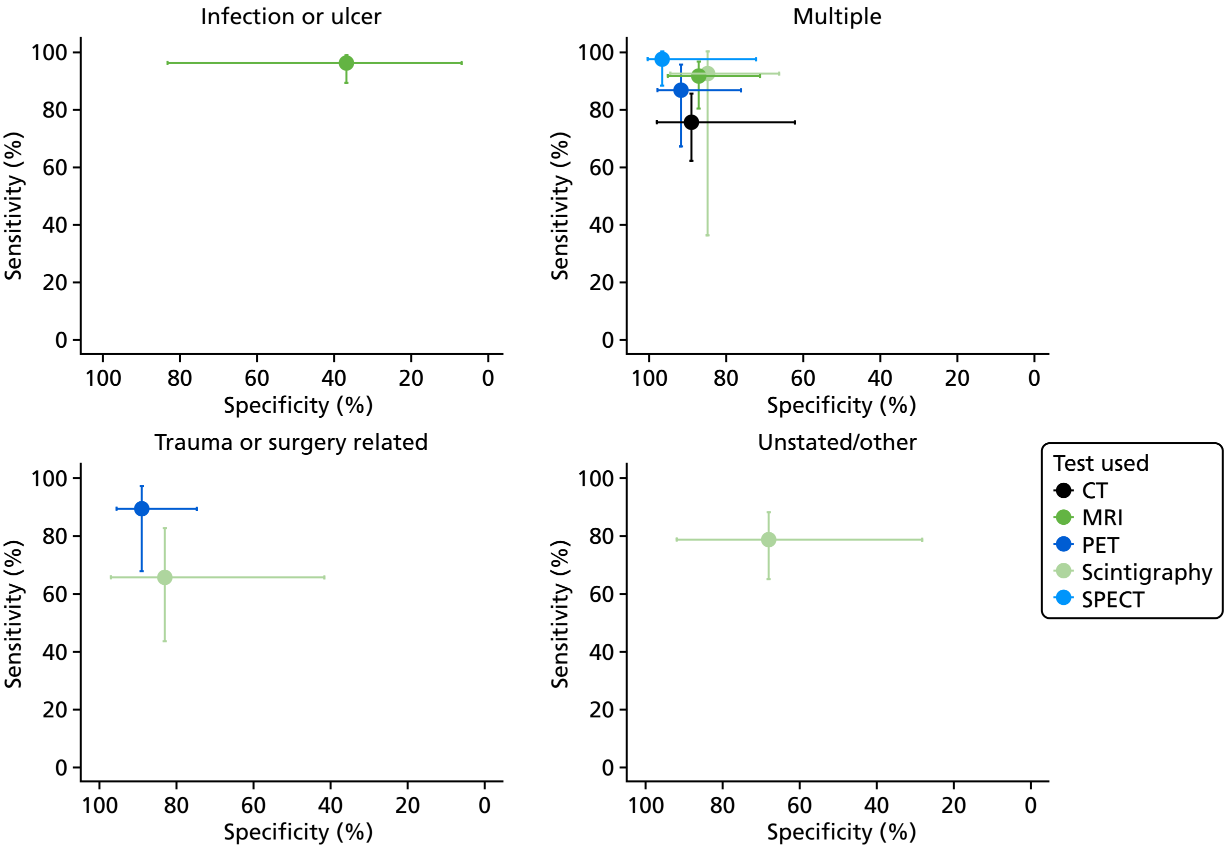


FIGURE 38 Summary sensitivity and specificity by cause of osteomyelitis.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

TABLE 25 Univariate meta-analysis – patients without diabetes

Test	Outcome	Estimate (%)	95% CI (%)	I^2
CT	Sensitivity	67.86	37.14 to 88.3	0
	Specificity	90.78	55.85 to 98.71	0
	DOR	9.74	3.33 to 28.48	10
	PR	51.72	32.82 to 81.51	20
MRI	Sensitivity	95.26	87.48 to 98.3	0
	Specificity	76.27	55.69 to 89.15	0
	DOR	30.28	10.64 to 86.19	43
	PR	80.69	71.97 to 90.47	22
PET	Sensitivity	80.75	59.87 to 92.18	0
	Specificity	90.25	67.8 to 97.6	0
	DOR	34.2	12.5 to 93.55	36
	PR	53.5	42.2 to 67.81	0
Scintigraphy	Sensitivity	83.01	65.12 to 92.74	0
	Specificity	69.32	48.95 to 84.19	0
	DOR	9.34	4.15 to 21.05	36
	PR	70.12	59.79 to 82.22	0
SPECT	Sensitivity	94.88	85.51 to 98.31	0
	Specificity	90.06	74.84 to 96.5	0
	DOR	95.82	21.89 to 419.35	52
	PR	72.32	61.11 to 85.58	0
Ultrasound	Sensitivity	96.77	52.46 to 99.88	–
	Specificity	30	0.49 to 97.37	–
	DOR	12.86	0.56 to 292.75	–
	PR	86.27	63.14 to 117.88	–
Radiography	Sensitivity	60.26	34.57 to 81.31	0
	Specificity	89.79	59.53 to 98.13	44
	DOR	7.07	2.71 to 18.41	0
	PR	56.29	40.33 to 78.57	0

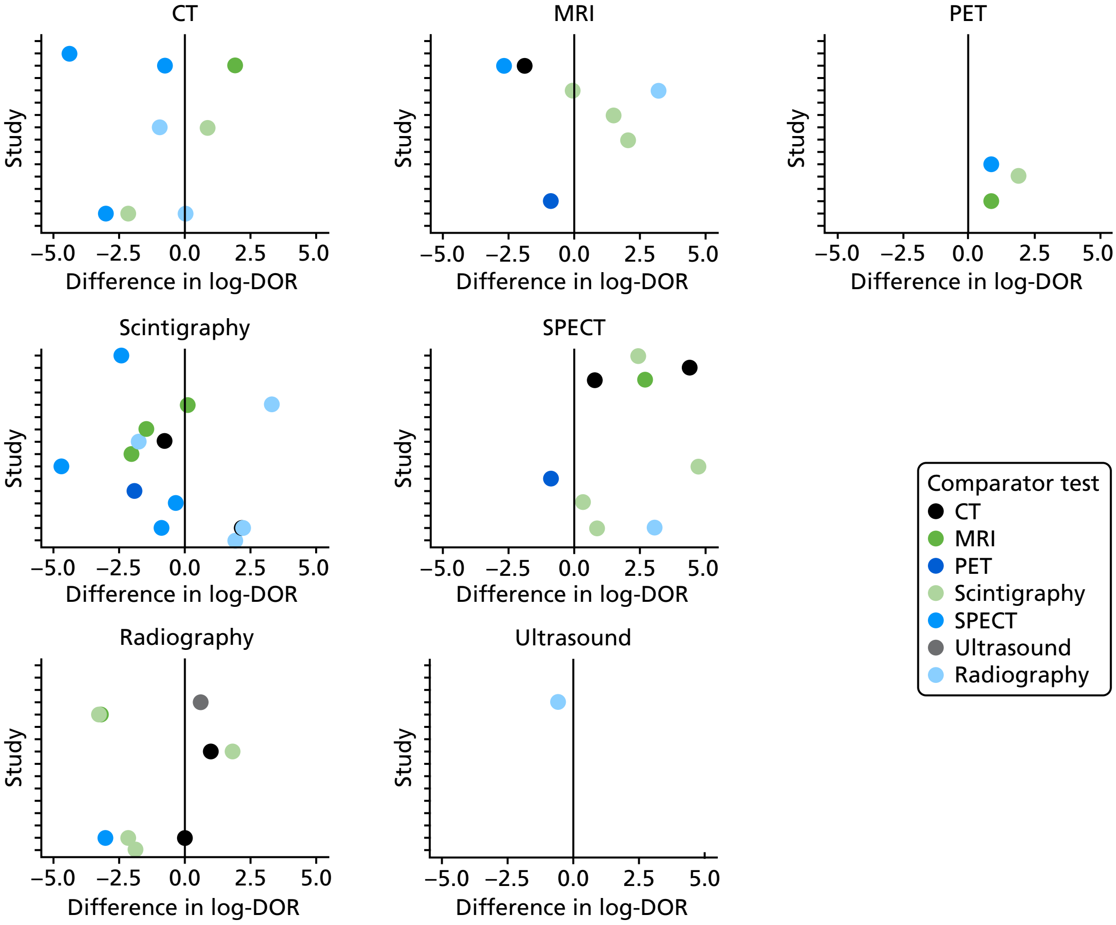


FIGURE 39 Comparisons of DORs across tests – patients without diabetes.

TABLE 26 Results of logistic regression to compare tests – patients without diabetes

	Difference in log-odds of specificity			Difference in log-DOR		
Test	**Estimate**	**Standard error**	***p*-value**	**Estimate**	**Standard error**	***p*-value**
Comparison with radiography						
MRI	0.472	0.598	0.43	2.062	0.898	0.022
Scintigraphy	0.989	0.412	0.016	0.325	0.567	0.566
CT	–0.144	0.545	0.792	0.813	0.687	0.236
SPECT	0.408	0.538	0.447	3.108	0.841	0
PET	–1.003	0.829	0.226	3.607	1.092	0.001
Comparison with MRI						
Scintigraphy	0.507	0.519	0.329	–1.722	0.797	0.031
Radiography	–0.484	0.598	0.418	–2.045	0.897	0.023
CT	–0.625	0.667	0.348	–1.236	0.946	0.191
SPECT	–0.074	0.581	0.899	1.061	0.942	0.26
PET	–1.488	0.811	0.066	1.562	1.076	0.147

© Queen's Printer and Controller of HMSO 2019. This work was produced by Llewellyn *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care

Published by the NIHR Journals Library