

Imaging tests for the detection of osteomyelitis: a systematic review

Alexis Llewellyn,¹ Julie Jones-Diette,¹
Jeannette Kraft,² Colin Holton,² Melissa Harden¹
and Mark Simmonds^{1*}

¹Centre for Reviews and Dissemination, University of York, York, UK

²Leeds Teaching Hospitals NHS Trust, Leeds, UK

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

Declared competing interests of authors: none

Published October 2019

DOI: 10.3310/hta23610

Scientific summary

Imaging tests for the detection of osteomyelitis

Health Technology Assessment 2019; Vol. 23: No. 61

DOI: 10.3310/hta23610

NIHR Journals Library www.journalslibrary.nihr.ac.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).

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.

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 that 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.

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.

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@nhr.ac.uk

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