What is the added value of ultrasound joint examination for monitoring synovitis in rheumatoid arthritis and can it be used to guide treatment decisions? A systematic review and cost-effectiveness analysis

Emma Simpson,[1]* Emma Hock,[1] Matt Stevenson,[1] Ruth Wong,[1] Naila Dracup,[1] Allan Wailoo,[1] Philip Conaghan,[2,3] Cristina Estrach,[4] Christopher Edwards[5] and Richard Wakefield[2,3]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
[3]National Institute for Health Research (NIHR) Leeds Biomedical Research Centre, Leeds, UK
[4]Aintree University Hospitals NHS Foundation Trust, Liverpool, UK
[5]National Institute for Health Research (NIHR) Wellcome Trust Clinical Research Facility, University of Southampton, Southampton, UK

*Corresponding author e.l.simpson@sheffield.ac.uk

Declared competing interests of authors: Richard Wakefield has provided consulting advice and spoken for General Electric with regard to ultrasound technologies and has also received speaker fees from AbbVie Inc. for ultrasound-related projects. Cristina Estrach has been a member of advisory boards for and/or received speaker fees from AbbVie Inc., Chugai Pharma (UK) Ltd and General Electric Co. Her institution has received educational grants from Pfizer Inc.

Published April 2018
DOI: 10.3310/hta22200

Scientific summary

Ultrasound joint examination for monitoring synovitis in rheumatoid arthritis

Health Technology Assessment 2018; Vol. 22: No. 20
DOI: 10.3310/hta22200

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

Scientific summary

Background

Rheumatoid arthritis (RA) is a chronic inflammatory disease characterised by progressive, irreversible joint damage, impaired joint function, pain and tenderness caused by swelling of the synovial lining of joints (synovitis) and is manifested by increasing disability and reduced quality of life. RA is associated with substantial costs both directly (associated with drug acquisition and hospitalisation) and indirectly, because of reduced productivity. Synovitis is assessed by clinical examination (CE) of the joints. Synovitis can also be assessed using imaging technologies including magnetic resonance imaging (MRI) and ultrasound (US); this may detect synovitis that is not detected by CE (subclinical synovitis) and may also distinguish between synovitis and other pathologies more readily than CE alone. US can therefore aid key decision-making with regard to therapy changes, leading to escalation or tapering of therapy. This is important for a disease area in which modern therapies are expensive and all therapies are associated with side effects, especially infections.

Objective

This report aimed to address the question: 'What is the added value of US joint examination for monitoring synovitis in RA and can it be used to guide treatment decisions?'

Data sources

The following electronic databases were searched: MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (via Ovid) (1946 to October 2015), EMBASE (via Ovid) (1974 to October 2015), Cochrane Database of Systematic Reviews (CDSR) (1996 to October 2015), Cochrane Central Register of Controlled Trials (CENTRAL) (1898 to October 2015), Health Technology Assessment (HTA) database (1989 to October 2015), Database of Abstracts of Reviews of Effects (DARE) (1946–2014), NHS Economic Evaluation Database (NHS EED) (1968–2014), Science Citation Index Expanded (1900 to October 2015), Science Citation Index and Conference Proceedings Index (1900 to October 2015), ClinicalTrials.gov (October 2015), European League Against Rheumatism Abstract Archive (via Web of Science) (October 2015), American College of Rheumatology and Association of Rheumatology Health Professionals (via Web of Science) (October 2015) and Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) conference proceedings (via Web of Science) (October 2015).

Review methods

A systematic review of US was conducted. Studies were sought that compared grey-scale US (GSUS) or power Doppler US (PDUS) with CE, the use of inflammatory biomarkers or disease activity scoring tools. The patient group considered was adults with RA. Outcomes included diagnostic (detection of synovitis, responsiveness to change), prognostic (association with progression or other disease outcomes) or treatment-related (response to treatment, treatment tapering, influence on decisions) measures. Diagnostic studies, prognostic studies and studies investigating the prediction of response to treatment, or treatment tapering, or the influence of US on treatment decisions were included. Although there are other methods for detecting synovitis, the decision was taken to limit the intervention in the review to US. Study selection was carried out by two reviewers, with disagreements resolved by discussion. Included studies were quality assessed and data extracted by one reviewer and checked by another reviewer. Data were tabulated and discussed. Study heterogeneity precluded meta-analyses.

A review of the cost-effectiveness of the use of US to monitor synovitis and a systematic search for the outcomes associated with tapering of RA treatment, irrespective of whether or not US was used, were undertaken. Study selection was carried out by two reviewers: any study deemed relevant by at least one reviewer was retrieved. Included studies were summarised by one reviewer. The tapering search was supplemented by checking the reference lists of included studies, searching for subsequent publications related to any abstracts identified and retrieving papers known to our clinical advisors. Relevant studies were summarised separately for cost-effectiveness and tapering. A survey publicised to UK rheumatology units was undertaken to investigate whether or not US is being used to monitor synovitis and guide treatment decisions in RA.

Mathematical model

The modelling undertaken was purposefully simplistic so that the key interactions between monitoring synovitis with US and decisions influencing treatment could be examined explicitly. The simple model estimated for patients in whom the clinician was contemplating reducing the treatment dose included (1) the reduction in treatment costs and therapy modification leading to serious infection avoidance that would be required for the addition of US for monitoring synovitis to have a cost per quality-adjusted life-year (QALY) gained of £20,000 and £30,000 and (2) the reduction in treatment costs required for the addition of US for monitoring synovitis to become cost saving. For patients in whom the clinician was thinking of increasing the treatment dose, two analyses were undertaken: (1) the reduction in the number of patients not progressing to more intensive treatment therapy or avoiding a serious infection through the use of US needed to achieve cost per QALY gained values of £20,000 and £30,000 and (2) the reduction in the number of patients escalating treatment needed for US to become cost saving.

Results

In the systematic review, 2724 records were identified from the electronic databases and an additional 26 records were identified from bibliography searching. Following title and abstract sifting, 154 articles were assessed for eligibility, of which 63 full-text papers were excluded. In total, 75 articles describing 58 studies were included; additionally, one study identified by the search as ongoing was published prior to publication of this report. A further 16 articles were retained for bibliography checking. Twenty-six studies provided prognostic and/or treatment data and 32 studies provided diagnostic data only.

Two randomised controlled trials (RCTs) of the treatment strategy did not find significant benefits in terms of the primary outcome of adding US to a Disease Activity Score (DAS)-based treat-to-target strategy for early RA patients. The addition of PDUS to a Disease Activity Score 28 joints-based treat-to-target strategy in the Targeting Synovitis in Early Rheumatoid Arthritis (TaSER) RCT resulted in no significant between-group difference in change from baseline in Disease Activity Score 44 joints (DAS44) (mean change: intervention –2.69, control –2.58). This study found that the addition of PDUS to the treatment strategy led to significantly more patients attaining DAS44 remission (66%) than for the DAS-alone strategy (43%) ($p = 0.03$). The Aiming for Remission in Rheumatoid Arthritis (ARCTIC) RCT found that the addition of PDUS and GSUS to a DAS-based strategy did not produce a significant between-group difference in the primary end point [which consisted of a composite of DAS of < 1.6, no swollen joints at 16, 20 and 24 months and no progression in van der Heijde-modified total Sharp score (vdHSS) between 16 and 24 months], with values of 22.0% and 18.8%, respectively. The ARCTIC trial did find that the change in erosion score of the vdHS had a significant advantage for the US group over the DAS group (changes of 0.5 and 1.0 for the US and control groups, respectively; $p = 0.04$). Erosion in the TaSER trial, as measured by change in the erosion score of the rheumatoid arthritis magnetic resonance imaging scoring system (RAMRIS), was not significantly different between the groups (changes of 0.5 and 1.0 for the US and control groups, respectively). These studies did not however explore the value of US when added only in cases of clinical uncertainty, for example, when there was discrepancy between DAS and clinical evaluation.

The majority of prospective cohort studies investigating radiographic progression reported that US at baseline, either GSUS or PDUS, was significantly correlated with progression at follow-up ($p = 0.05$ to $p < 0.001$). Radiographic progression in most, but not all, cases was measured with a modification of the total Sharp score/vdHSS. PDUS was significantly associated with radiographic progression in 10 studies in which it was measured. Associations were reported as odds ratios (ORs) in two studies and were 12.21 [95% confidence interval (CI) 3.34 to 44.73; $p < 0.001$] and 1.80 (95% CI 1.20 to 2.71; $p = 0.005$). Associations reported as correlation coefficients ranged from $r = 0.099$ to $r = 0.77$ ($p = 0.05$ to $p < 0.001$). Significance levels from Mann–Whitney *U*-test results reported were $p < 0.001$ and $p = 0.0011$. GSUS was significantly associated with radiographic progression in six studies but not in three studies. Significant associations reported were Mann–Whitney *U*-test $p = 0.027$; $r = 0.140$ to $r = 0.61$ ($p < 0.001$); and ORs of 2.08 (95% CI 1.39 to 3.11) and 2.15 (95% CI 1.23 to 3.75) ($p = 0.01$). The difference between studies reporting significant associations and studies reporting non-significant associations could not be explained by study quality, joints assessed or how the end point was measured.

Other outcomes reported were heterogeneous, making it difficult to draw conclusions. PDUS was significantly correlated with the proportion of patients experiencing disease flare at follow-up ($p = 0.014$), whereas this was not significant for GSUS. US could significantly predict treatment persistence ($p = 0.02$) and US predicted treatment tapering or discontinuation failure significantly in two out of three studies ($p = 0.005$, $p < 0.0005$, $p = 0.06$) whereas clinical measures alone did not. The additional use of US modified treatment decisions (in 23–88% of cases in UK studies) and significantly increased ($p < 0.001$ to $p < 0.0005$) clinician confidence in treatment decisions.

The review of cost-effectiveness studies identified five articles, although none was directly relevant to the decision problem. Nineteen papers were identified from the tapering search, which, when supplemented by checking references of identified articles, by searching for subsequent articles related to identified abstracts and by articles known to our clinical advisors, resulted in 39 relevant papers being included. Given that evidence showed that some patients who had achieved a low level of disease activity could have their treatment tapered, with no or little short-term harm to them, it was deemed appropriate to model strategies in which a clinician was contemplating a dose reduction. The survey conducted yielded only 31 responses, 27 of which stated that US was used for treatment decisions. The small sample size means that the results can not be generalised across the UK.

For patients who have been stable on biological disease-modifying anti-rheumatic drugs (bDMARDs) and in whom the clinician is contemplating reducing the dose of bDMARDs, the model estimated that an average reduction of 2.5% in the costs of bDMARDs was sufficient to cover the costs of performing US every 3 months. Similarly, if 2.5% of patients do not have their treatment escalated to bDMARDs, it was estimated that the use of US to monitor synovitis would be cost neutral. If only conventional disease-modifying anti-rheumatic drugs (DMARDs) were considered in the current or planned treatment regimen, the money spent on US monitoring could not be recouped.

Limitations

Few RCTs were available and so lower-quality study designs were included in the review. The heterogeneity of the studies identified precluded meta-analysis. Therefore, no summary estimates of effect were available, which is a limitation of the review. There is no gold standard/reference standard for the detection of synovitis, although it has been suggested that MRI may be used as a gold standard/reference standard; it detects similar levels of inflammation to US. The systematic search for DMARD tapering was not overly sensitive as many articles were identified by bibliography and citation searching and contact with clinical advisors (rather than by database searching). However, the summarised articles provide a broad overview of the literature base and it is likely that any papers omitted would not have an adverse impact on the findings of this report.

Limitations in the modelling include the exclusion of biosimilar bDMARDs, which are likely to reduce the cost associated with bDMARD treatment compared with the base case. However, a sensitivity analyses was conducted to explore the impact of price reductions on the threshold levels.

A key limitation within the modelling was the lack of robust data relating to key parameters within the decision problem. As such, threshold analyses have been undertaken to provide indicative levels of drug dose reduction and avoidance of serious infections required to make the use of US cost neutral.

Conclusion

Limited evidence was available and therefore cost-effectiveness analysis was limited to threshold analysis. Given the proportions of patients who could potentially taper treatment, or remain on stable therapy without escalation, the use of US to monitor synovitis could potentially be a cost-effective approach if it provides clinicians with more confidence in reducing the drug burden. However, there is considerable uncertainty in this conclusion.

The most important future research studies would be longitudinal studies evaluating the role of US in the management of synovitis in RA.

Study registration

This study is registered as PROSPERO CRD42015017216.

Funding

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

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

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

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

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.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 14/16/01. The contractual start date was in January 2015. The draft report began editorial review in April 2016 and was accepted for publication in August 2017. 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.

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