Total ankle replacement versus ankle arthrodesis for patients aged 50–85 years with end-stage ankle osteoarthritis: the TARVA RCT

Andrew J Goldberg,[1,2,3*] Kashfia Chowdhury,[4] Ekaterina Bordea,[4] James Blackstone,[4] Deirdre Brooking,[2] Elizabeth L Deane,[4] Iva Hauptmannova,[2] Paul Cooke,[5] Marion Cumbers,[2] Simon S Skene[6] and Caroline J Doré[4] on behalf of the TARVA Study Group[†]

[1]Institute of Orthopaedics & Musculoskeletal Science, Division of Surgery, University College London, London, UK
[2]Department of Research & Innovation, Royal National Orthopaedic Hospital, London, UK
[3]MSK Lab, Sir Michael Uren Hub, Imperial College London, London, UK
[4]Comprehensive Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK
[5]Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Trust, Oxford, UK
[6]Surrey Clinical Trials Unit, University of Surrey, Guildford, UK

*Corresponding author Andy.goldberg@nhs.net
[†]The names of the TARVA Study Group members can be found at www.anklearthritis.co.uk

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/PTYJ1146.

Primary conflicts of interest: Andrew J Goldberg declares the following competing interests: American Orthopaedic Foot and Ankle Society (board or committee member; 2012–present), British Orthopaedic Foot & Ankle Society (board or committee member; 2011–present), National Joint Registry (board or committee member; 2010–present), Foot and Ankle International and Foot and Ankle Orthopaedics (editorial or governing board; 2014–present), Stryker (previously Wright Medical Technology, Inc.; Memphis, TN, USA) (paid presenter or speaker), Standing CT (stock or stock options) and X-Bolt Orthopaedics (Dublin, Ireland) (unpaid consultant)..

Published March 2023
DOI: 10.3310/PTYJ1146

Scientific summary

Total ankle replacement versus ankle arthrodesis

Health Technology Assessment 2023; Vol. 27: No. 5
DOI: 10.3310/PTYJ1146

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

Scientific summary

Background

More than 29,000 patients in the UK present to specialists each year with symptomatic end-stage ankle osteoarthritis, a condition in which the cartilage lining the ankle joint has worn away, causing progressive pain and stiffness. Ankle osteoarthritis causes major disability and has a similar impact on quality of life (QoL) as end-stage hip osteoarthritis and cardiac failure. The demand incidence for ankle osteoarthritis has been estimated at 47.7 per 100,000. The majority of this is secondary to trauma caused by fractures or severe sprains, both of which are increasing; hence, ankle osteoarthritis is likely to become an increasingly important health problem, especially among working adults. Other causes of ankle osteoarthritis include long-standing inflammatory arthropathies (e.g. rheumatoid arthritis, haemochromatosis and haemophiliac arthropathy).

In the early stages of disease, non-operative measures such as a change in activity levels, weight loss, physiotherapy, painkillers and ankle braces should be used. When these conservative measures have failed, and a surgeon confirms the diagnosis of end-stage osteoarthritis on the basis of radiological and clinical evidence (i.e. plain radiographs and unrelenting symptoms, respectively), surgery might then be considered.

Although arthrodesis (i.e. ankle fusion) is the most common surgical treatment for end-stage ankle osteoarthritis, in response to patient demand, more and more surgeons are performing total ankle replacement (TAR). At least 4000 patients are treated with ankle fusion or TAR each year in the NHS. The TARs implanted in England, Wales, Northern Ireland, the Isle of Man and Guernsey are captured on the National Joint Registry, which has revision surgery as its end point. The British Orthopaedic Foot & Ankle Society only recently started capturing outcome data on ankle fusion patients. All studies comparing TAR with ankle fusion to date have been observational and, to the best of our knowledge, there have been no high-quality prospective randomised trials reported.

Objectives

The total ankle replacement versus ankle arthrodesis (TARVA) trial aimed to compare the clinical effectiveness and cost-effectiveness of TAR with that of ankle fusion for the treatment of end-stage ankle osteoarthritis in patients aged 50–85 years. Clinical effectiveness was measured through self-reported pain-free function using a standardised questionnaire of walking and standing ability 52 weeks post surgery. We also aimed to determine whether or not there was a difference in physical function [measured using the Foot and Ankle Ability Measure – Activities of Daily Living (FAAM-ADL)], QoL [measured using the EuroQol 5-Dimensions, five-level version (EQ-5D-5L)] and range of ankle motion at 26 and 52 weeks post surgery. We investigated the cost-effectiveness and cost–utility of TAR and ankle fusion.

Methods

Design

We conducted a pragmatic prospective, multicentre, parallel-group, non-blinded randomised controlled trial (RCT). Participants were randomised equally between two arms: TAR and ankle fusion. The study protocol [Goldberg AJ, Zaidi R, Thomson C, Doré CJ, Skene SS, Cro S, *et al.* Total ankle replacement versus arthrodesis (TARVA): protocol for a multicentre randomised controlled trial. *BMJ Open* 2016;**6**:e012716] was developed before recruitment commenced and detailed the design, interventions and study procedures.

Setting and participants

The trial was conducted across 17 participating UK sites. Patients with end-stage ankle osteoarthritis who were aged 50–85 years and who the surgeon believed were suitable for both TAR and ankle fusion were eligible to join the trial. Patients had to be able to read and understand the patient information sheet (PIS) and provide written consent on an informed consent form (ICF).

Interventions and follow-up

At randomisation patients were allocated to receive either TAR or ankle fusion. For TAR, the remaining damaged cartilage was removed and the joints resurfaced with metal implants and an intervening polyethylene liner, either fixed or mobile bearing, to act as a gliding surface. All prostheses were Conformité Européenne marked. For ankle fusion, the remaining damaged cartilage was removed from the ends of the bone and the two bones held together in compression using screws or plates to join them as one bone (bone fusion) so that there was no longer any movement at the tibiotalar joint. Participants were seen at recruitment, randomisation, surgery visit and at 2, 6, 12, 26 and 52 weeks post surgery.

Blinding

This was an open (non-blinded) trial. It was not possible to blind patients, surgeons, radiologists or clinical assessors.

Randomisation

Participants were randomised in a 1 : 1 ratio to either TAR or ankle fusion. Randomisation was carried out using minimisation incorporating a random element, with surgeon and whether osteoarthritis was present in the subtalar or the talonavicular joint as minimisation factors. A secure online service (Sealed Envelope™; Sealed Envelope Ltd, London, UK) provided the treatment arm allocations.

Outcome measures

The primary outcome was the change in the Manchester–Oxford Foot Questionnaire (MOXFQ) walking/standing domain scores (0–100, where low scores are better) between the preoperative baseline and 52 weeks post surgery. The secondary outcomes were change in MOXFQ walking/standing domain score from preoperative baseline to 26 weeks and change in MOXFQ pain and social interaction domain scores from baseline to 26 and 52 weeks. An additional measure of physical function, the Foot and Ankle Ability Measure (FAAM), was captured at baseline and at 26 and 52 weeks. The changes in FAAM-ADL (0–100, where higher scores are better) and FAAM sport subscale scores from baseline were compared between arms. We also compared changes in QoL from baseline to 12, 26 and 52 weeks using the EQ-5D-5L questionnaire. Longer-term follow-up at 2, 5 and 10 years post surgery is planned.

Total range of motion (ROM) of the tibia to the floor was captured at baseline and 52 weeks. All adverse events (AEs), serious adverse events (SAEs) and complications reported from the date of consent until 52 weeks were compared between arms. Secondary outcomes that related to the economic evaluation included quality-adjusted life-years (QALYs), health-care resource use collected from patient files and a modified version of the Client Service Receipt Inventory (CSRI), and mean incremental cost per QALY gained.

Sample size

The sample size calculation for the primary outcome (change in MOXFQ walking/standing domain score by 52 weeks) was performed using Stata/IC®, version 12.1 (StataCorp LP, College Station, TX, USA). It was based on achieving 90% power to detect an estimated minimal important difference (MID) in the primary outcome at the 5% level of significance, accounting for expected loss to follow-up.

The sample size calculation was based on previous observational studies and determined it was important to detect a difference of 12 in the change from baseline in the MOXFQ walking/standing domain score between the two treatment arms. The standard deviation (SD) of the MOXFQ walking/standing domain score was estimated to be 27, and loss to follow-up was estimated to be 10% (attrition in similar RCTs has been 5–7%). Based on these quantities, the required sample size was estimated to be 118 patients per arm. The sample size was adjusted to account for clustering by surgeon. The intraclass

correlation coefficient (ICC) was estimated based on previous studies, and the initially computed sample size was inflated by a factor of $f = 1 + (m - 1) \times ICC$. Assuming an average cluster size (m) of 14 (patients per surgeon) and an ICC of 0.03, an inflation factor of $f = 1.39$ was estimated, leading to a final required sample size of 164 per arm or 328 patients in total.

Data collection and management

Data were entered into a central MACRO v4 database (Elsevier, Amsterdam, the Netherlands) by sites, with internal validation checks to improve data quality; data queries were resolved by site staff before database lock and final analysis.

Statistical methods

As per the statistical analysis plan, all the analyses were conducted on an intention-to-treat (ITT) basis, meaning that all randomised participants with at least one postsurgery follow-up visit were included in the analysis, regardless of their adherence to treatment. In addition, a per-protocol analysis was performed for the primary outcome that included outcome data from only those patients who received their randomised surgical procedure within the time window specified in the protocol.

The primary analysis involved fitting a multilevel repeated-measures linear regression model to estimate the difference between treatment arms in the change in MOXFQ walking/standing domain score from baseline to 52 weeks. This analysis model used all available visit data (from 26 weeks and 52 weeks) to strengthen confidence in the missing at random assumption and provide greater power to detect differences at individual visits. The model was adjusted for baseline MOXFQ walking/standing domain score and presence of osteoarthritis in each of the two adjacent joints. A random surgeon effect was also included in the model to account for clustering by surgeon. Similar models were used for other continuous secondary outcomes to estimate differences at 26 and 52 weeks post surgery.

Economic evaluation

The aim of the economic evaluation was to assess the cost-effectiveness of TAR compared with ankle fusion in patients with end-stage osteoarthritis. We compared the costs and outcomes of the two arms over the time horizon of 52 weeks. Outcomes were QALYs, calculated using utility index values obtained from the EQ-5D-5L. The primary within-trial analysis was conducted according to the ITT principle from the NHS and personal social services (PSS) perspective. Costs included cost of surgery, cost of health-care resource use (collected using the CSRI) and cost of concomitant medications. Sensitivity analyses included per-protocol analysis and analysis from a societal perspective. The societal perspective included additional out-of-pocket costs incurred by the participants and any productivity loss. The analytical approach is a cost–utility analysis as it estimates the mean incremental cost per QALY gained of TAR compared with ankle fusion.

We built a decision model to extrapolate the trial results to a lifetime horizon. We constructed a simple Markov model, which simulates participants' pathways after TAR or ankle fusion. Monte Carlo simulations were used to account for uncertainty. We estimated the probability of the intervention being cost-effective at the cost-effectiveness thresholds of £20,000–30,000 per QALY gained, recommended by the National Institute for Health and Care Excellence (NICE).

Results

Baseline characteristics

Between March 2015 and January 2019, 303 participants were randomised; 282 participants had surgery and 281 attended at least one follow-up visit. The mean age was 68 years, 71% of participants were male and 43% had arthritis in one or more adjacent joints. The arms were well balanced at baseline, as observed from the baseline characteristics.

Primary outcome

The mean (SD) MOXFQ walking/standing domain score at 52 weeks was 31.4 (30.4) in the TAR arm and 36.8 (30.6) in the ankle fusion arm. Patients improved in both arms, but the adjusted mean [95%

confidence interval (CI)] difference of −5.56 (−12.49 to 1.37) suggests that, on average, patients who received TAR had a MOXFQ walking/standing score 5.56 points lower than those who received ankle fusion at 52 weeks post surgery. This difference was not statistically significant ($p = 0.12$). The 95% CI included the MID of −12, so the trial was not able to exclude the MID. After 52 weeks, more patients achieved the MID in the TAR arm (82%) than in the ankle fusion arm (80%).

In a post hoc analysis, when each of the two TAR subtypes (fixed- and mobile-bearing implants) were compared with ankle fusion, the mean (SD) MOXFQ walking/standing domain score at 52 weeks was 25.9 (28.3) in the fixed-bearing TAR arm and 36.8 (30.6) in the ankle fusion arm. The adjusted difference of −11.1 (95% CI −19.3 to −2.9) suggests that, on average, patients who received a fixed-bearing TAR had a MOXFQ walking/standing score 11.1 points lower than those who received ankle fusion at 52 weeks post surgery. This difference was statistically significant ($p = 0.008$).

Secondary outcomes

The MOXFQ pain and social interaction domain scores also suggested improvement in patients in both arms, but the adjusted difference of −4.20 (95% CI −9.80 to 1.39) for pain and −5.06 (95% CI −10.37 to 0.26) for social interaction at 52 weeks post surgery were not statistically significant ($p = 0.14$ and $p = 0.06$, respectively). The difference between the TAR and ankle fusion arms in the change in MOXFQ walking/standing domain score at 26 weeks was statistically significant ($p = 0.02$).

The difference between the TAR and ankle fusion arms in the change in FAAM-ADL scores at 52 weeks was statistically significant ($p = 0.01$). There were improvements from baseline in both arms, but a difference of 6.16 (95% CI 1.54 to 10.78) between arms. The change in the EQ-5D-5L visual analogue scale value was statistically significant at 26 weeks ($p = 0.03$), but the change in the EQ-5D-5L index value was not significantly different at 26 weeks ($p = 0.08$) and 52 weeks ($p = 0.32$) between the two treatment arms.

At 52 weeks from baseline, the ROM (dorsiflexion and plantarflexion) improved for patients with TAR and decreased for those with ankle fusion; the difference between arms was statistically significant ($p < 0.001$). One or more SAE occurred in 17.8% of TAR and 23.8% of ankle fusion patients ($p = 0.19$). One or more AE occurred in 54.3% of TAR and 52.6% of ankle fusion patients ($p = 0.84$). The risks of patients experiencing any SAE or AE during the course of the trial were not statistically significantly different between the two arms.

Economic evaluation

Total ankle replacement generated more QALYs than ankle fusion, but this difference was not statistically significant (adjusted difference 0.02, 95% CI −0.008 to 0.05; $p = 0.14$). The CI was generated using the bootstrapping technique (1000 iterations). The total cost of TAR from the NHS and PSS perspective was £2576 higher than the total cost of ankle fusion (95% CI £1181 to £3988; $p < 0.01$). The difference was due to the difference in the cost of surgery (£2230, 95% CI £1024 to £3103; $p < 0.01$), as other differences in other cost components were not statistically significant. The incremental cost-effectiveness ratio (ICER) was £127,931 per QALY gained at 52 weeks.

Model-based analysis suggested that TAR is cost-saving compared with ankle fusion when extrapolated to a lifetime horizon. As the population of interest is aged 50–85 years, the average life expectancy was 17 years; therefore, the model was run for 17 cycles. Over the lifetime horizon, there was a 69% probability that TAR would be cost-effective compared with ankle fusion at the cost-effectiveness threshold of £20,000 per QALY gained.

Conclusions

Both TAR and ankle fusion improved patients' QoL at 1 year, but we did not show one to be superior in terms of clinical scores at 52 weeks when using either ITT or per-protocol analysis. The TARVA trial is inconclusive in terms of the superiority of TAR, as the 95% CI for the adjusted treatment effect includes both a difference of zero and the MID of 12. However, we can rule out the superiority of ankle fusion.

Both operations appear to be safe. A post hoc analysis of the most common type of implant in the UK, the fixed-bearing TAR, did show a statistically significant improvement of TAR over ankle fusion, suggesting that fixed-bearing TAR may outperform ankle fusion. There is a 69% probability of TAR being cost-effective compared with ankle fusion at the NICE cost-effectiveness threshold £20,000 per QALY gained over patients' lifetime.

Future research

There is a strong case for continuing follow-up, in particular to study the radiological and clinical progress of these patients, and the need for revision surgery. There is also a need for studies to explore the sensitivity of clinically important differences between arms when both have already improved significantly from their baseline scores.

Trial registration

This trial is registered as ISRCTN60672307 and ClinicalTrials.gov NCT02128555.

Funding

This project was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 27, No. 5. See the NIHR Journals Library website for further project information.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4.014 and is ranked 27th (out of 108 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2021 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), Embase (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded (Clarivate , Philadelphia, PA, USA).

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.

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 16/78/04. The contractual start date was in May 2018. The draft report began editorial review in November 2020 and was accepted for publication in May 2021. 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 and Care 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, 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, the HTA programme or the Department of Health and Social Care.

Published by NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress, final files produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).