Unicompartmental compared with total knee replacement for patients with multimorbidities: a cohort study using propensity score stratification and inverse probability weighting

Albert Prats-Uribe,[1] Spyros Kolovos,[1] Klara Berencsi,[1] Andrew Carr,[1] Andrew Judge,[1,2] Alan Silman,[1] Nigel Arden,[1,3,4] Irene Petersen,[5] Ian J Douglas,[6] J Mark Wilkinson,[7,8] David Murray,[1] Jose M Valderas,[9] David J Beard,[1] Sarah E Lamb,[1,10] M Sanni Ali,[1,6] Rafael Pinedo-Villanueva,[1] Victoria Y Strauss[1*] and Daniel Prieto-Alhambra[1]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Nuffield Orthopaedic Centre, University of Oxford, Oxford, UK
[2]Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, University of Bristol, Southmead Hospital, Bristol, UK
[3]Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, Botnar Research Centre, Nuffield Orthopaedic Centre, University of Oxford, Oxford, UK
[4]Medical Research Council Lifecourse Epidemiological Unit, University of Southampton, Southampton, UK
[5]Department of Primary Care and Population Health, University College London, London, UK
[6]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[7]Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK
[8]Research Committee, National Joint Registry for England, Wales, Northern Ireland and the Isle of Man, Hemel Hempstead, UK
[9]College of Medicine and Health, University of Exeter, Exeter, UK
[10]University of Exeter Medical School, Institute of Health Research, College of Medicine and Health, Exeter, UK

*Corresponding author Victoria.strauss@csm.ox.ac.uk

Declared competing interests of authors: Albert Prats-Uribe reports grants from Fundación Alfonso Martin Escudero (Madrid, Spain) and the Medical Research Council (London, UK). Andrew Judge was a sub-panel member of the National Institute for Health Research (NIHR) Programme Grants for Applied Research (PGfAR) programme (September 2015–August 2020); has received personal fees from Freshfields Bruckhaus Deringer (London, UK); and was a member of the Data Safety and Monitoring Board (December 2012 to June 2016) (which involved receipt of fees) from Anthera Pharmaceuticals Inc. (Hayward, CA, USA). Nigel Arden reports grants from Merck & Co. Inc. (Darmstadt, Germany) and

personal fees from Merck & Co. Inc., Flexion Therapeutics (Burlington, MA, USA), Regeneron (Tarrytown, NY, USA) and Pfizer Inc. (New York, NY, USA)/Eli Lilly and Company (Indianapolis, IN, USA) outside the submitted work. David Murray reports research grants from NIHR Health Technology Assessment (HTA), specifically for TOPKAT (HTA 08/14/08), outside the submitted work; institutional grant and consultancy fees to the University of Oxford with Zimmer Biomet (Warsaw, IN, USA); and royalties from knee replacement-related patents. David J Beard reports grants from the NIHR HTA programme during the conduct of the study, specifically for TOPKAT. Sarah E Lamb was on the NIHR HTA Additional Capacity Funding Board (2012–15), HTA End of Life Care and Add-on Studies Board (2015), HTA Prioritisation Group Board (2010–15) and the HTA Trauma Board (2007–8). Rafael Pinedo-Villanueva reports personal fees from Mereo BioPharma Group plc (London, UK), Kyowa Kirin International (Galashiels, UK) and UCB Biopharma SPRL (Brussels, Belgium) outside the submitted work. Daniel Prieto-Alhambra reports NIHR HTA Funding Committee membership (November 2017–present); research grants from Amgen Inc. (Thousand Oaks, CA, USA), Johnson & Johnson (New Brunswick, NJ, USA), and UCB Biopharma SPRL; speaker services and advisory board membership fees paid to his department/research group from Amgen; and consultancy fees paid to his department/research group from UCB Biopharma SPRL. He also reports that Janssen Pharmaceutica (Beerse, Belgium), on behalf of Innovative Medicines Initiative-funded European Health Data Evidence Network and European Medical Information Framework consortiums, and Synapse Management Partners have supported internal training programmes organised by his department and open training programmes for external participants.

Published November 2021
DOI: 10.3310/hta25660

Scientific summary

Unicompartmental compared with total knee replacement

Health Technology Assessment 2021; Vol. 25: No. 66
DOI: 10.3310/hta25660

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

Scientific summary

Background

Routinely collected NHS clinical data and national registries offer new opportunities for the comparative assessment of health technologies in actual practice conditions. This is particularly interesting for elderly and complex patients with multiple comorbidities, who are excluded from many randomised controlled trials. Surgical randomised controlled trials are particularly challenging owing to ethics difficulties, scarce surgeon equipoise and the need for specialised and experienced treatment centres and teams.

Two procedures for knee arthroplasty that are offered in the NHS (unicompartmental and total knee replacement) were compared in a National Institute for Health Research Health Technology Assessment programme-funded surgical randomised controlled trial [08/14/08; Total or Partial Knee Arthroplasty Trial (TOPKAT)]. Although TOPKAT offered top-quality information on the comparative effects of these surgeries for relatively healthy (American Society of Anesthesiologists grade of 1 or 2) patients, data from the National Joint Registry suggest that almost one in six patients undergoing unicompartmental or total knee replacement surgery in the UK have an American Society of Anesthesiologists grade of ≥ 3. The TOPKAT findings are, thus, hard to interpret for a substantial proportion of NHS patients.

Routinely collected data contain information on, potentially, all NHS patients, regardless of their medical history. These data sets offer an opportunity for research that includes elderly and multimorbid participants. However, the lack of random allocation of treatments in such databases does pose challenges, including confounding by indication. If confounding is not accounted for and minimised, it can lead to bias.

Objectives

In stage 1 of the Unicompartmental (vs. Total) knee replacement for patients with Multimorbidity Study (UTMoSt), we assessed whether or not the available analytical methods could obtain comparable findings to those from TOPKAT, using participants in the National Joint Registry who would have been eligible for TOPKAT (American Society of Anesthesiologists grade of 1 or 2). The proposed propensity score and instrumental variable methods were each applied to the data set. Those offering results comparable to TOPKAT were deemed valid and were used in stage 2.

In stage 2 of UTMoSt, the validated methods from stage 1 were used to compare the benefits (postoperative patient-reported outcome measures), risks (revision, complications and mortality), hospital costs and cost-effectiveness of unicompartmental and total knee replacement among National Joint Registry participants who would not have been eligible for TOPKAT (American Society of Anesthesiologists grade of ≥ 3).

Methods

For data sources, National Joint Registry participants undergoing total or unicompartmental knee replacement with linked, routinely collected data from the NHS hospital inpatient records were included in safety analyses. Those with linked patient-reported outcome measure data were included in the primary outcome analyses.

The participants in stage 1 were total and unicompartmental knee replacement recipients recorded in the National Joint Registry with linked data who would have been eligible for TOPKAT. In stage 2, participants were recruited who had an American Society of Anesthesiologists grade of 3 or 4, indicating severe systemic comorbidities that would have excluded them from TOPKAT. The comparison was unicompartmental versus total knee replacement.

The primary outcome was postoperative Oxford Knee Score (patient-reported outcome measure). The secondary outcomes were safety outcomes, including 90-day risks of venous thromboembolism, myocardial infarction and prosthetic joint infection (stage 2 only), and 5-year risks of revision and mortality. The health economic analysis outcomes were health-related quality of life (EuroQol-5 Dimensions) and NHS hospital costs (stage 2 only).

Statistics

In stage 1, four propensity score-based approaches and inverse probability weighting were used to account for measured confounding: (1) propensity score matching (1 : 5), (2) stratification based on the distribution of the propensity score in the whole cohort, (3) stratification based on the unicompartmental knee replacement cohort and (4) propensity score adjustment (linear and non-linear models). For each outcome, a logistic regression model was used to calculate the propensity score for unicompartmental knee replacement using patient-level characteristics, including demographics, preoperative patient-reported outcome measures, comorbidities and procedures recorded within the 3 years before surgery. Missing body mass index data and preoperative patient-reported outcome measures were imputed using multiple imputation by chained equations. Covariate balance was assessed using absolute standardised mean difference, with a predefined cut-off point of 0.1.

We also explored four potential instrumental variables: surgeon preference, hospital preference, geographical location and calendar time. When certain assumptions are fulfilled, instrumental variable analyses can account for measured and unmeasured confounders. Key instrumental variable assumptions were checked with *F*-statistics, odds ratios (strength of the instrument) and absolute standardised mean differences (lack of an association between the instrument and the confounders).

We compared the results obtained for each method with the TOPKAT findings using the TOPKAT outcome analysis methods: multilevel linear regression for postoperative Oxford Knee Score and a multilevel Poisson model for 5-year revision or death. Two-stage analyses were used for instrumental variables. We predefined three criteria by which an analytical method would be considered unable to replicate the TOPKAT findings and, therefore, be invalid for stage 2: chi-squared test p-value < 0.05, a relatively large tau^2 or an $I^2 > 40\%$. We also used two newly proposed methods to assess the methods' validity: whether or not the obtained treatment effect estimates fall within the trial's 95% confidence interval and statistical significance agreement. We performed sensitivity analyses on the valid methods, including restricting the analysis to surgeries performed by lead surgeons with ≥ 10, ≥ 30 and ≥ 50 index surgeries in the previous year.

In stage 2, for each valid method and each outcome, patient-level characteristics overall and for unicompartmental knee replacement patients were compared using absolute standardised mean difference with a cut-off point of 0.1. Differences in postoperative Oxford Knee Score between unicompartmental knee replacement patients and total knee replacement patients were estimated using multilevel linear regression. For each of the 90-day postoperative complications, the relative risk and 95% confidence interval were estimated using Poisson models with robust standard errors. Cause-specific hazard models were fitted to estimate the risk of 5-year revision or mortality, censoring patients when they had revision or mortality (a competing event). Prespecified interactions between surgery types and sex, age or American Society of Anesthesiologists grade were assessed with a

p-value of < 0.1. Long-term complications were also assessed when restricting the analysis to patients operated on by experienced surgeons.

For the health economic evaluation, multilevel regression analyses were performed to estimate the differences in costs and quality-adjusted life-years between unicompartmental knee replacement and total knee replacement patients. The regression models for quality-adjusted life-years also included the preoperative utility score as a covariate. The incremental cost-effectiveness ratio was calculated by dividing the difference in costs by the difference in quality-adjusted life-years between unicompartmental knee replacement and total knee replacement patients. The uncertainty surrounding the incremental cost-effectiveness ratio was estimated using non-parametric bootstrapping with 1000 replications.

Results

In stage 1, 21,026 National Joint Registry participants undergoing unicompartmental knee replacement and 273,530 participants undergoing total knee replacement would have been eligible for TOPKAT. Of these participants, 1197 unicompartmental knee replacement and 125,834 total knee replacement patients had postoperative Oxford Knee Score data and could be included in the Oxford Knee Score analysis.

In the Oxford Knee Score analysis, inverse probability weighting and propensity score stratification based on the whole cohort resulted in unresolved imbalances, whereas propensity score matching and propensity score stratification based on the unicompartmental knee replacement cohort achieved good balance. All of the propensity score-based methods resulted in an average treatment effect estimate favouring unicompartmental knee replacement, but with at least 1 point less than the effect seen in the trial, ranging from 0.10 (propensity score non-linear adjustment) to 0.76 (propensity score stratification based on the unicompartmental knee replacement cohort), compared with 1.91 in TOPKAT.

Propensity score stratification based on the unicompartmental knee replacement cohort was the preferred method ($I^2 = 35\%$, chi-squared test $p = 0.21$ and $\tau^2 = 0.23$), followed by inverse probability weighting ($I^2 = 48\%$, chi-squared test $p = 0.17$ and $\tau^2 = 0.43$) and propensity score stratification based on the whole cohort ($I^2 = 53\%$, chi-squared test $p = 0.14$ and $\tau^2 = 0.48$).

A surgeon-level eligibility criterion was then applied to mimic surgeon eligibility in TOPKAT, including only participants operated on by surgeons who had performed ≥ 10 surgeries of the same type in the previous year. The treatment estimates from all three methods moved closer to the TOPKAT findings, with average treatment effects of 1.37 (95% confidence interval 0.54 to 2.20) for propensity score stratification based on the unicompartmental knee replacement cohort, 1.37 (95% confidence interval 0.54 to 2.20) for propensity score stratification based on the whole cohort and 1.32 (95% confidence interval 0.32 to 2.33) for inverse probability weighting, compared with 1.91 (95% confidence interval 0.20 to 3.62) in TOPKAT. All three methods had an I^2 of 0% and small τ^2, indicating that they were able to replicate TOPKAT findings.

Only five of the potential instrumental variables passed both testable assumptions: the three lead surgeon-based preference instruments (based on 20, 30 and 50 previous surgeries) and two of the consultant surgeon-based preference instruments (based on 30 and 50 previous surgeries). The other tested instrumental variables violated either one or both of the testable assumptions. The five selected instrumental variables then all failed to produce a comparable treatment effect estimate with TOPKAT, with a chi-squared test *p*-value < 0.001 and $I^2 > 90\%$. All of the instrumental variable analyses passed the statistical significance agreement tests and showed a significant improvement in postoperative Oxford Knee Score favouring unicompartmental knee replacement, as in TOPKAT.

In stage 2, the comparative safety analyses included 57,682 total knee replacement patients and 2256 unicompartmental knee replacement patients. Of these patients, only 145 unicompartmental knee replacement and 23,344 total knee replacement patients were included in the Oxford Knee Score analysis. Propensity score stratification based on the unicompartmental knee replacement cohort yielded excellent covariate balance both between and within strata. Propensity score stratification based on the whole cohort had excellent average covariate balance between the 10 strata. Four covariates remained imbalanced after inverse probability weighting. Propensity score stratification based on the unicompartmental knee replacement cohort and on the whole cohort resulted in statistically significant positive effects for unicompartmental knee replacement, with an estimated mean postoperative Oxford Knee Score difference of 1.83 (95% confidence interval 0.10 to 3.56) points and 1.82 (95% confidence interval 0.10 to 3.56) points in favour of unicompartmental knee replacement, respectively, which is close to the effect seen in TOPKAT. Inverse probability weighting analysis found an insignificant effect in postoperative Oxford Knee Score.

Unicompartmental knee replacement patients had a lower relative risk of developing venous thromboembolism in the 90 days after surgery than total knee replacement patients, with relative risks of 0.33 (95% confidence interval 0.15 to 0.74) based on propensity score stratification and 0.39 (95% confidence interval 0.16 to 0.96) based on inverse probability weighting. No significant differences in myocardial infarction or prosthetic joint infection risks were found between unicompartmental knee replacement and total knee replacement patients. Unicompartmental knee replacement patients experienced a higher risk of revision over 5 years than total knee replacement patients, with hazard ratios of 2.70 (95% confidence interval 2.15 to 3.38) in propensity score stratification analyses and 2.60 (95% confidence interval 1.94 to 3.97) in inverse probability weighting. They also had reduced all-cause mortality in propensity score stratification analyses, with a hazard ratio of 0.52 (95% confidence interval 0.36 to 0.74). However, this difference was attenuated when using inverse probability weighting.

American Society of Anesthesiologists grade and sex had significant interactions with total knee replacement and unicompartmental knee replacement: women had a higher risk of revision than men, and people with an American Society of Anesthesiologists grade of 4 had a much higher revision risk than patients with an American Society of Anesthesiologists grade of 3, although statistical power was a concern.

The crude mean cost of a primary knee replacement was £6246 (standard deviation £779) for unicompartmental knee replacement patients and £6627 (standard deviation £1402) for total knee replacement patients. The mean costs for complications were £3560 (standard deviation £6) for unicompartmental knee replacement patients and £3986 (standard deviation £3853) for total knee replacement patients. The mean differences in quality-adjusted life-years gained were 0.147 (95% confidence interval −0.507 to 0.803) and 0.330 (95% confidence interval −0.305 to 0.967) in favour of unicompartmental knee replacement when using inverse probability weighting and propensity score stratification, respectively. Unicompartmental knee replacement costs were £334 (95% confidence interval £306 to £362) and £359 (95% confidence interval £339 to £378) lower than total knee replacement costs, using inverse probability weighting and propensity score stratification, respectively.

Conclusions

Propensity score-based stratification and inverse probability weighting successfully replicated the TOPKAT findings in the primary outcome (postoperative Oxford Knee Score) analyses, indicating that these methods can be used to minimise confounding in observational studies on the comparative effectiveness of implantable medical devices. Propensity score adjustment, propensity score matching and instrumental variable methods led to results that departed from those observed in TOPKAT. More research is required on the best use of analytical methods and design of observational post-marketing research of medical devices.

In stage 2, unicompartmental knee replacement had similar effectiveness for patients with multimorbidity as for the healthier (stage 1 and TOPKAT) population. There was little or no clinically relevant difference in postoperative Oxford Knee Score between unicompartmental knee replacement and total knee replacement patients. A strongly protective effect against postoperative venous thromboembolism for patients undergoing unicompartmental knee replacement was identified. In the long term, unicompartmental knee replacement was associated with an almost threefold higher revision risk than total knee replacement, but also with a reduction in all-cause mortality of almost 50%. Cost-effectiveness analyses showed that unicompartmental knee replacement dominated in patients with substantial comorbidity (American Society of Anesthesiologists grade of ≥ 3), as it was both more beneficial and less expensive than the alternative (total knee replacement) in this patient subgroup. These findings should guide future clinical guidelines on knee replacement for patients with severe multimorbidity.

Trial registration

This trial is registered as EUPAS17435.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 66. 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

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and 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 15/80/40. The contractual start date was in June 2017. The draft report began editorial review in January 2020 and was accepted for publication in July 2020. 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).