

The Lancet Rheumatology

How long does a shoulder replacement last? A systematic review and meta-analysis of case-series and national registry reports with more than 10 years of follow-up --Manuscript Draft--

Manuscript Number:	TLRHEU-D-20-00161R1
Article Type:	Article (Meta-analysis)
Keywords:	Shoulder replacement; Shoulder arthroplasty; survival; Patient Reported Outcome Measures; PROMs
Corresponding Author:	Jonathan P Evans, MBChB, MD, FRCS University of Exeter Medical School UNITED KINGDOM
First Author:	Jonathan P Evans, MD
Order of Authors:	Jonathan P Evans, MD Jonathan T Evans, MD Richard S Craig, FRCS (T+O) Hasan R Mohammad, MRes Adrian Sayers, MSc Ashley W Blom, PhD Michael R Whitehouse, PhD Jonathan L Rees, MD
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	<p>Background Shoulder replacement is an increasingly common treatment for end-stage degenerative shoulder conditions. Some shoulder replacements will fail and further operations may be required. It is important for patients and clinicians to know how long shoulder replacements last and how effectively they improve pain and function. This study aims to determine the longevity and long-term efficacy of shoulder replacements.</p> <p>Methods In this systematic review and meta-analysis, we searched MEDLINE and Embase for articles reporting 10-year or greater survival of Total Shoulder Replacements (TSR), Humeral Hemiarthroplasties (HA) and Reverse Total Shoulder Replacements (RTSR). Survival, implant and Patient Reported Outcome Measures (PROMs) data were extracted. National joint replacement registries were reviewed and analysed separately. We weighted each series and calculated a pooled survival estimate at 10, 15 and 20 years. For PROMs we pooled the Standardised Mean Difference (SMD) at 10 years.</p> <p>Findings We identified 10 series reporting all-cause survival of 529 TSRs and 420 HA, no series for RTSR met our inclusion criteria. The estimated 10-year survival for TSR was 95·6% (95% CI 93·6, 97·6) and HA 90·4% (95% CI 87·0, 94·0). A single registry contributed 7941 TSRs, 3495 HAs and 8049 RTSRs. The pooled registry 10-year survival for TSR was 92·0% (95% CI 91·0, 93·0), HA 90·5% (95% CI 81·8, 95·1) and RTSR 94·4% (95% CI 93·1, 95·7) for osteoarthritis and 93·6% (95% CI 91·0, 95·4) for rotator cuff arthropathy. Pooled 10-year PROMs revealed a substantial improvement from baseline scores (SMD 2·13 95% CI 1·93, 2·34).</p> <p>Interpretation Over 90% of shoulder replacements last more than 10 years and patient reported benefits are sustained. This long overdue information will be of use to patients and health-care providers.</p> <p>Funding The National Institute for Health Research, the National Joint Registry for England, Wales, Northern Ireland, and Isle of Man, and the Royal College of Surgeons of England.</p>

How long does a shoulder replacement last? A systematic review and meta-analysis of case-series and national registry reports with more than 10 years of follow-up

Jonathan P Evans ^{MD}^{1,2}, Jonathan T Evans ^{MD}³, Richard S Craig ^{FRCS (T&O)}^{4,5}, Hasan R Mohammad ^{MRes}^{4,5},
Adrian Sayers ^{MSc}³, Prof Ashley W Blom ^{PhD}^{3,6}, Michael R Whitehouse ^{PhD}^{3,6}, Prof Jonathan L Rees ^{MD}^{4,5}

1. Health & Policy Research Group, University of Exeter, Exeter, EX1 2LU, UK
2. National Institute for Health Research Applied Research Collaboration South West Peninsula, University of Exeter, Exeter, EX1 2LU, UK
3. Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, 1st Floor Learning & Research Building, Southmead Hospital, Bristol, BS10 5NB, UK
4. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, OX3 7LD, UK
5. National Institute for Health Research Oxford Biomedical Research Centre, Oxford, OX3 7LD, UK
6. National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, BS10 5NB, UK

Correspondence to:

Mr Jonathan Peter Evans

Health & Policy Research Group, University of Exeter, Exeter, EX1 2LU, UK

j.p.evans2@exeter.ac.uk

Panel: Research in context

Evidence before this study

Survival of shoulder replacements has often been reported in small case-series, with some follow-up extending beyond 20 years, however individual case-series are prone to bias and reporting has been highly heterogeneous. We searched MEDLINE and Embase for systematic reviews and meta-analyses of shoulder replacement series that were published in English. Of the 37 systematic reviews we identified, no articles reported combined survival estimates or patient reported outcome measures with more than 10 years follow-up. A previous analysis of the UK Hospital Episode Statistics (HES) dataset, published in 2019, combined all types of shoulder implants and found overall survival to be 90.0% (95% CI 89.6% to 90.3%) at 10 years. No study to date has attempted to provide pooled survival estimates and pooled patient reported outcomes for shoulder replacements more than 10 years after surgery.

Added value of this study

To our knowledge, we provide the first pooled survival estimate, drawn from multiple sources, for shoulder replacements at 10 years. We have also shown that shoulder replacements have a sustained positive impact on patients' lives to 10 years after surgery. Our findings showed that approximately 92% of total shoulder replacements, 91% of shoulder humeral hemiarthroplasties and 94% of reverse total shoulder replacements last for 10 years.

Implications of all the available evidence

Our findings provide valuable and overdue information for patients and clinicians considering shoulder replacement surgery. It is the first study to provide a simple and generalizable answer to two very important questions: "How long does a shoulder replacement last?" and "Will my shoulder be better in the long-term after surgery?" The data will also be useful for those commissioning healthcare services.

55 Abstract

56 Background

57 Shoulder replacement is an increasingly common treatment for end-stage degenerative shoulder conditions.
58 Some shoulder replacements will fail and further operations may be required. It is important for patients
59 and clinicians to know how long shoulder replacements last and how effectively they improve pain and
60 function. This study aims to determine the longevity and long-term efficacy of shoulder replacements.

61 Methods

62 In this systematic review and meta-analysis, we searched MEDLINE and Embase for articles reporting 10-
63 year or greater survival of Total Shoulder Replacements (TSR), Humeral Hemiarthroplasties (HA) and
64 Reverse Total Shoulder Replacements (RTSR). Survival, implant and Patient Reported Outcome Measures
65 (PROMs) data were extracted. National joint replacement registries were reviewed and analysed separately.
66 We weighted each series and calculated a pooled survival estimate at 10, 15 and 20 years. For PROMs we
67 pooled the Standardised Mean Difference (SMD) at 10 years.

68 Findings

69 We identified 10 series reporting all-cause survival of 529 TSRs and 420 HA, no series for RTSR met our
70 inclusion criteria. The estimated 10-year survival for TSR was 95.6% (95% CI 93.6, 97.6) and HA 90.4%
71 (95% CI 87.0, 94.0). A single registry contributed 7941 TSRs, 3495 HAs and 8049 RTSRs. The pooled
72 registry 10-year survival for TSR was 92.0% (95% CI 91.0, 93.0), HA 90.5% (95% CI 81.8, 95.1) and
73 RTSR 94.4% (95% CI 93.1, 95.7) for osteoarthritis and 93.6% (95% CI 91.0, 95.4) for rotator cuff
74 arthropathy. Pooled 10-year PROMs revealed a substantial improvement from baseline scores (SMD 2.13
75 95% CI 1.93, 2.34).

76 Interpretation

77 Over 90% of shoulder replacements last more than 10 years and patient reported benefits are sustained.
78 This long overdue information will be of use to patients and health-care providers.

79 Funding

80 The National Institute for Health Research, the National Joint Registry for England, Wales, Northern
81 Ireland, and Isle of Man, and the Royal College of Surgeons of England.

Introduction

Patients with severe pain and disability from degenerative shoulder conditions want to know whether they will benefit from shoulder replacement surgery, which type of replacement may be best and what they can expect in the long-term following surgery.¹ A review of seven national arthroplasty registers in 2017 suggested there has been a secular increase in the number of shoulder replacements performed for patients with both osteoarthritis and rotator cuff tear arthropathy. Overall the annual incidence rate has increased 2.8 fold in the last decade, but significant variation exists between countries.² There is a paucity of high quality outcome data to aid joint decision making by patients and clinicians, and to assist both commissioners and providers in understanding the utility and likely revision burden associated with undertaking these procedures.

Available randomised controlled trials (RCTs) are particularly limited, by size and design, in their ability to evaluate the longer-term outcomes and risks of primary shoulder arthroplasty, in particular the requirement for revision surgery.³ To better understand the long-term benefits and risks of shoulder replacement surgery for these patients, a formal appraisal and synthesis of the more frequently available non-randomised study data is needed.

Ideally, clinicians and surgeons should be able to provide patients with contemporary condition-, age- and implant-specific outcome data for any proposed procedure and available alternatives. While implant manufacturers do facilitate the collection of implant-level data in order to gain relevant benchmark accreditation,⁴ detailed and reliable data are not yet available for shoulders. Until such granular brand-level information is available, clinicians and patients need accurate information on classes of available implants. Hip and knee replacement have shown that although there is variation between brands, classes of implants behave in broadly similar fashion.^{5,6} The three main constructs or classes available and referred to in this study are conventional total shoulder replacement (TSR), humeral hemiarthroplasty (HA), and reverse total shoulder replacement (RTSR). There is likely to be heterogeneity between indications for surgery, mechanisms of failure and overall revision rates between these different constructs.⁷

107 In this study we sought to answer a simple but important question posed by all patients: How long does a
108 shoulder replacement last? We aimed to provide the best quality pooled estimates of implant survival at a
109 minimum 10 years' follow-up. The decision to revise a poorly performing shoulder replacement is
110 multifactorial that may be sensitive to both patient and surgeon preferences. Therefore, we also aimed to
111 make a pooled estimate of the likely patient reported outcome at long-term follow-up, in essence to answer
112 the question: Will my shoulder be better 10 years after surgery?

113

Methods

Search strategy and selection criteria

We conducted a systematic review and meta-analysis assessing the survival of shoulder replacements in case-series and national joint registries following a predefined protocol registered with PROSPERO (CRD42019140221) and complying with PRISMA guidelines.⁸

A search strategy using keywords and MeSH terms relating to shoulder replacement and survival (appendix 1) was used in the databases MEDLINE and Embase accessed through OVID Silver Platter. The databases were searched from their commencement to 24th September 2019. The strategy development was guided by previously published search strategies exploring the survival of hip and knee replacements.^{9,10} Manual screening of the bibliographies of the full-text articles and systematic reviews was also undertaken.

Studies were included if they assessed patients who had undergone any type of shoulder replacement (a total shoulder replacement (TSR), humeral hemiarthroplasty (HA) or reverse total shoulder replacement (RTSR)). Humeral components (stemmed, stemless or resurfacing) were all considered as TSR or HA dependent on whether the glenoid (shoulder socket) was replaced or not and not sub-classified. The indication (reason) for surgery had to be predominantly osteoarthritis (OA) or rotator cuff arthropathy (RTCA). For inclusion, the case-series or published registry report had to report the survival of a specific brand of implant with a mean or median follow-up of greater than 10 years. It is widely accepted that survival of hip arthroplasties varies by the brand of implant.⁵ Although this has not specifically been assessed in shoulder replacements, the technique of treating each brand as its own series was utilised as variation in survival by brand exists in hip and knee replacements, therefore the assumption would seem sensible for shoulder replacements as well. Weighting of implants in the meta-analysis would therefore provide the most robust survival estimates. This allows us to treat each series as an individual study and weight the meta-analysis of survival results according to the standard error of each series. Aggregate data from multiple implant brands would not allow this granularity and thus hide the potential variability in performance between implant brands. A cut-off of minimum mean or median follow-up of 10 years was

chosen as the subject of interest of this study was “long-term” survival, where there is a current paucity of information. We accept this definition may vary subjectively but 10 years allowed inclusion of sufficient studies to make analyses robust and represents a time period that is relatable to patients and clinicians.

Studies were excluded if they reported the outcome of revision surgery, as this is often more complex surgery and carries different survivorship. Conference abstracts were excluded due to the limited data available from these reports. Systematic reviews were assessed for their citations but did not include their pooled data to avoid duplication.

The reports from all available national joint registers that collect and publish the individual implant-specific survivorship for shoulder replacements with at least 10-years of follow-up were assessed. Reports were identified through the systematic search if published or accessed through their websites.

Article screening and data extraction

Screening was undertaken in a stepwise manner using the web application Rayyan.¹¹ Journal article titles and abstracts were screened by two reviewers (JTE and HM) with arbitration of conflict undertaken by JPE. Full-text review and data extraction were undertaken by two reviewers independently (JPE and JTE). Data extracted were: publication date, baseline population demographics, number of patients (n), surgical indication proportion (% OA and/or % RCTA), follow-up duration (>10 years), implant name and construct type (TSA, HA or RTSA), loss to follow-up, survival estimates (including CIs) and all available Patient Reported Outcome Measure (PROM) (e.g. Visual Analogue Scales (VAS), Constant score, Disabilities of the Arm, Shoulder and Hand (DASH)), data (outcome measure used baseline mean score (SD), follow-up duration in 5 year increments, follow-up mean score (SD)). Data were not extracted from figures (e.g. Kaplan Meier plots) to avoid potential transcription inaccuracy. Discrepancy in extracted data was discussed by the authors, following which there were no cases of disagreement.

Statistical Analysis

For the assessment of the published case-series our primary exposure was the shoulder replacement implant and our primary outcome was all-cause revision, of any part of this construct, as guided by our patient

group.¹² Statistical analysis was performed with Stata 15 (*Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC). Survival estimates, assuming that survivorship approximated revision risk, were pooled by meta-analysis. Each series was weighted according to its standard error (calculated from published confidence intervals). The effect size (Standardised Mean Difference (SMD)) of the primary PROMs reported in each study was pooled with meta-analysis with weighting according to sample size and analysed using a random effects model as a more conservative estimate of treatment effect. Effect size was considered small if it was less than ≥ 0.2 , moderate if ≥ 0.5 and large if ≥ 0.8 .¹³

Quality assessment

Study quality was assessed using the non-summative four-point system (consecutive cases, multi-centre, under 20% loss to follow-up and use of multivariable analysis) developed by Wylde et al.¹⁴ This was selected in preference to the summative MINORS score due to the high loss to follow-up in joint replacement case-series and because some of the scoring criteria in MINORS were not relevant to joint replacement.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had access to the raw data. The corresponding author had full access to all of the data and the final responsibility to submit for publication

Results

The search of published case-series produced 1,376 articles. Of these, 449 duplicates were removed, leaving 927 articles for screening (figure 1). After screening, 36 full-text articles were reviewed. Additional citation searches through previously published systematic review references yielded four further full-text reviews, none of which met the inclusion criteria. Following review of full-text articles, nine articles reporting 10 individual implant specific series were included in the survival analysis, six articles that reported both survival analysis and PROMs were included in the PROMs analysis. A summary of study level characteristics is provided in Table 1. The proportion of OA as the primary surgical indication was 59% for TSR and 48% for HA. The reporting of indication was variable and was interpretable in only seven articles. Quality assessment revealed that six (60%) of the 10 series were consecutive, two (20%) were multicenter, nine (90%) had >80% follow-up (with mean loss to follow up of 8.4%, ranging from 0% to 23.7%), and none undertook multivariable analysis. These proportions are in keeping with the fact that the quality of published case-series is low.

Case-series

Six unique series, published between 1998 – 2015, reported survival of 529 total shoulder replacements (TSR) at 13 time points with follow-up ranging from 10 to 21 years (Appendix 2).^{15–21} Four reported survival at exactly 10 years (466 TSRs), three reported survival at 15 years (427 TSRs) and one reported survival at 20 years (19 TSRs). Pooled survival from those studies reporting at exactly 10 years was 95.6% (95% CI 93.6, 97.6) at 15 years 88.5% (95% CI 83.4, 94.1) and at 20 years 83.2% (95% CI 70.5, 97.8) (figure 2). When studies reported survival estimates at between 10 and 15 years, these results were rounded down to 10 years as a sensitivity analysis. This resulted in a pooled survival of six series (529 TSRs) of 90.0% (95% CI 88.3, 91.7) (figure 3).

Four unique series, published between 1998 – 2017, reported survival of 364 shoulder humeral hemiarthroplasties (HAs) at 10 time points with follow-up ranging from 10 to 21 years (Appendix 2).^{16,18,21–23} Three reported survival at exactly 10 years (327 HAs), two at 15 years (151 HAs) and one at 20 years

(56 HAs). Pooled survival at exactly 10 years was 90.4% (95% CI 87.0, 94.0), at 15 years 90.6% (95% CI 84.1, 97.1), and at 20 years 75.6% (95% CI 65.9, 86.5) (figure 2). Rounding down of reported survival from those series closest to >10 but <15 years resulted in a pooled survival of four series (364 HAs) of 92.5% (95% CI 89.6, 95.3) (figure 3).

No unique single implant series with a mean follow-up of at least 10 years were found for reverse total shoulder replacements (RTSA).

Registry data

The reports of implant-level data at 10 years were only available from a single registry, the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) 2019 annual report.²⁴ This report yielded 10-year survival of eight series of TSRs (7,941 arthroplasties), eight series of HAs (3,495 arthroplasties) and five series of RTSRs (8,049 arthroplasties). Pooled survival estimates from registry data for TSRs at 10 years were 92.0% (95% CI 91.0, 93.0); for HAs 90.5% (95% CI 81.8, 95.1) and for RTSRs were 94.4% (95% CI 93.1, 95.7) for a primary diagnosis of OA, and 93.6% (95% CI 91.0, 95.4) for a diagnosis of RTCA (single implant reported) (figure 4).

Patient Reported Outcome Measures

Of the 14 studies reporting survival analysis, six reported the implant level PROMs of 617 shoulder replacements for inclusion in the PROMs meta-analysis; this included two studies not included in the survival meta-analysis, excluded as they did not report confidence intervals.^{17,19,20,23,25,26} Four studies reported PROMs on TSR, one on RTSR and one on HA. All reported the outcome of shoulder-specific PROMs, without the addition of generic quality of life measures. Five studies reported the Constant score, one the simple shoulder test (SST) and one a four-point linear pain scale previously described by Neer.²⁷ Pooled PROMs data showed a large effect of improved outcome from baseline (SMD 2.13 95% CI 1.93, 2.34) (figure 5). Subgroup analysis of PROMs exclusively from TSRs reduced the effect size marginally (SMD 2.02 95% CI 1.86, 2.19). Implant-level 10-year PROMs were not published in any registry reports. The New Zealand registry report 10-year PROMs, which were categorised by construct only (TSR, HA,

232 RTSR, Partial resurfacing of head). Although no baseline PROMs are available for comparison, at 10-years
233 the Oxford Shoulder Score (OSS) mean for all implants was 39.1/48 (95% CI 38.4, 39.8), for TSA (n=335)
234 41.0/48 (40.0, 42.0), HA (n=104) 39.4/48 (37.7, 41.1), RTSR (n=104) 39.4 (37.7, 41.1).

Discussion

We found that 90% of shoulder replacements last for at least 10 years and that patients can expect a large and sustained improvement in their patient reported outcome measures.

The methodology used is one that has been previously applied successfully to hip and knee replacement,^{9,10} with the production of simple and generalisable results. The application of this process to shoulder replacement proved more complex due to sparsity and heterogeneity of data and highlights why the study question has not previously been answered. However, despite these limitations, the data from both registries and case-series independently estimate the same results. This is encouraging and suggests that these case-series are not subject to selection and publication bias.

The methods applied in this study use an individual estimate for each implant series, which is then synthesised to provide single pooled construct estimate weighted according to the standard error.. The implant has been shown to be fundamental to the survival outcome of hip and knee replacement and is likely to be just as important in shoulder replacement and each individual series should be considered as a different patient cohort.⁵ We have used the individual estimates for each implant to synthesise a single pooled estimate, weighting the estimates according to standard error. This type of analysis, deriving an overall estimate according to how frequently each implant has been used, is unique to our study. This analysis is dependent upon case-series, and registries' reporting of implant level data, as the only method where the patterns of implant failure can be accounted for. .

Implant survival at more than 10 years was greater than 90% for both TSR and HA in the case-series data, and also in the Australian registry data. This finding is concordant with the limited number of extended survival reports using multi-implant cohorts, including the assessment of Hospital Episode Statistics (HES) data in England²⁸ of 90% (95%CI 89·6 - 90·3) in a combined arthroplasty cohort, and Mayo clinic registry data^{29,30} of 90·2% (95% CI 88·7, 91·7) for TSR and 90·0% (95% CI 88·0, 92·0) for HA. This study found very limited extended case-series 20-year data, all from the Mayo group, with survival for TSRs of 83·2% and HAs 75·6%, which are lower than the HES report of 87·8% (95% CI 87·2, 88·4) at 18 years but

comparable to the full Mayo Clinic registry of 81.4% (95% CI 78.4, 84.5) for TSR, but worse than the HA survival of 85.0% (95% CI 81.8, 88.4) at a 20 years, notably there is a younger age cohort in their HA case-series. It is notable that the demographic characteristics from the case-series and registry data are similar for the TSR group, and concordantly their survival rates are also comparable. For the HA group, the case-series data contain a more male dominated and younger population. All but one of the case-series report an average age of <60yrs, therefore the survival findings from case-series may lack generalisability.

For RTSR, there was an absence of any implant level data from case-series at more than 10 years. This is concerning as it is currently utilized in over 50% of shoulder replacements in the UK, Norway, Australia and New Zealand.^{24,31-33} It is surprising that this change in practice has occurred so rapidly with such paucity of long-term outcome evidence, particularly after the well documented problems with the widespread adoption of unproven technology in joint replacement.⁶ It is therefore reassuring that we have been able to assess survival of RTSR at 10 years using data synthesised from the Australian registry data which reveals a survival of 94.0% (95% CI 93.1, 95.7) for OA and 93.6% (95% CI 91.0, 95.4) for RTCA.

Of the studies that reported survival of shoulder replacements at a mean of >10 years, five did not include confidence intervals and could not be added to the meta-analysis, six reported the composite survival of cohorts that included multiple different implants. Addition of these data would have resulted in the inclusion of 1,482 arthroplasties, increasing the analysis cohort by >150%. Failure of individual components of the construct (e.g. the glenoid or humeral component in isolation) was also reported in a large series that was excluded from the meta-analysis owing to the absence of an all-cause construct survival estimate.³⁴ Although component-failure data are of interest, we would regard this as best reported as a secondary endpoint, with the all-cause 1-Kaplan Meier estimate as the most appropriate method of reporting survivorship, which should always include the number of shoulder replacements at risk at the time of reporting.³⁵

As shoulder replacement registries may not provide long-term survival for some time to come, we remain somewhat reliant on case-series data. If these series are to reliably inform the surgical community of

implants at risk, they must be transparently reported according to current guidance on the reporting of healthcare data.³⁶ As novel implants and techniques are developed, we will also continue to be reliant upon case-series to highlight potential improvements in survivorship and function.

This study has identified that at over 10 years from the primary intervention a large improvement (SMD 2.13) in PROMs scores was maintained. A linear transformation, making all scores interpretable from the Constant score scale, also demonstrates a mean change score of 40.4, which exceeds the minimal clinically reported difference (MCID) of 12.8 ± 2.5 points for TSR.³⁷ The authors recognise the concern regarding the validity of the Constant score, and suggest that future studies report PROMs with proven validity and responsiveness. The New Zealand registry provided the only published comparator of construct-level, but not implant level, PROMs data. At 10 years this was limited to 674 replacements. Their high OSS at 10 years (80% of total score) does suggest a sustained benefit of shoulder replacements. As the New Zealand registry does not provide baseline pre-operative scores, comparison of SMD could not be undertaken.

We echo the calls for consensus in outcome choice to facilitate synthesis of data. Initiatives that promote the use of core outcome sets include the Core Outcome Measures in Effectiveness Trials (COMET), Outcome Measures in Rheumatology (OMERACT) and the International Consortium for Health Outcome Measurement (ICHOM).³⁸⁻⁴⁰ Furthermore, the inclusion of PROMs in registry data has the potential to dramatically improve the assessment of patient-focused outcomes. Currently, clear associations between survival of a shoulder implant and the patient-focused domains of pain, function and quality of life cannot be ascertained.

There are limitations of this work. The data did not allow stratification or adjustment for patient factors that may have affected outcomes in the pooled analysis. The analysis could not account for differing thresholds for revision between surgeons. It is notable that many of the historic series are derived from single-surgeon series and therefore surgeon preferences may alter the resultant weighted synthesis of survivorship. We also recognise that emergent techniques and implants may demonstrate superior (or inferior) survivorship and function that is yet to be demonstrated with long-term follow-up. The impact of historic series that have

310 utilised implants subsequently recognised as having worse outcomes can affect a synthesis of long-term
311 outcomes. The series from Levy et al ¹⁶ which included metal-backed glenoid components had a large
312 weighting that reduced the overall survival estimate. Reporting early failure of certain implants is important
313 and for the best available overall estimates should continue to be included. As not all failure results in
314 revision, we reported patient-reported outcomes to better define the overall value of shoulder replacement.
315 Our pooled registry results are drawn exclusively from the Australian register. As the available follow-up
316 in other registries increases, a wealth of data will soon become available, and we would encourage implant
317 level reporting by brand and product line. We also assumed that survival estimates are equivalent to risks
318 for generating pooled estimates, and although the assumption that no censoring occurs (patients dying with
319 a shoulder in situ) is violated, it provides a useful method of aggregation in the absence of individual patient
320 data. The aggregated estimates of survival are however the largest possible sample and this is the largest
321 report of this type and length of follow-up.

322 The strengths of this study include an inclusive and comprehensive design and realistic interpretation of
323 survivorship that accounts for all revisions and not a limited or biased subset, as well as a patient outcome
324 focus. From a patient perspective, all revision surgery carries risk and therefore all-cause revision should
325 be considered.

326 Conclusion

327 By pooling survival from case-series and registry data, we have been able to provide a reliable estimate of
328 10-year survival rate of shoulder replacements. We found that over 90% of shoulder replacements last for
329 at least 10 years. Patients experienced sustained and marked benefit to 10 years. This information should
330 be reassuring for patients, health professionals and commissioners of health services.

331

Contributions

JPE and JTE were responsible for study concept, design, screening, data extraction, data analysis, and writing of this manuscript.

HM completed the primary screening of abstracts and review of the manuscript.

JR, AB, MRW, RC and AS were responsible for study concept, design, and writing of the manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgments

This report is independent research supported by the National Institute for Health Research Applied Research Collaboration South West Peninsula, National Institute for Health Research Biomedical Research Centre at the University Hospitals Bristol National Health Service (NHS) Foundation Trust and the University of Bristol, and National Institute for Health Research Oxford Biomedical Research Centre. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Funding

This study was supported by the NIHR Applied Research Collaboration South West Peninsula, NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and University of Bristol, and NIHR Oxford Biomedical Research Centre.

JPE was supported by an NIHR Clinical Lectureship.

JTE was supported by the joint National Joint Registry of England, Wales, Northern Ireland and the Isle of Man and Royal College of Surgeons of England Fellowship.

AS was supported by an MRC Strategic Skills Fellowship MR/L01226X/1

References

- 1 Rangan A, Upadhaya S, Regan S, Toye F, Rees JL. Research priorities for shoulder surgery: Results of the 2015 James Lind Alliance patient and clinician priority setting partnership. *BMJ Open* 2016; **6**: e010412.
- 2 Lübbecke A, Rees JL, Barea C, Combescure C, Carr AJ, Silman AJ. International variation in shoulder arthroplasty. *Acta Orthop* 2017; **88**: 592–9.
- 3 Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL. Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy. *Cochrane Database Syst Rev* 2020.
- 4 Orthopaedic Data Evaluation Panel (ODEP). <http://www.odep.org.uk/> (accessed March 11, 2020).
- 5 Deere KC, Whitehouse MR, Porter M, Blom AW, Sayers A. Assessing the non-inferiority of prosthesis constructs used in hip replacement using data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of man: a benchmarking study. *BMJ Open* 2019; **9**: e026685.
- 6 Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet* 2012; **379**: 1199–204.
- 7 Lapner PLC, Rollins MD, Netting C, Tuna M, Bader Eddeen A, van Walraven C. A population-based comparison of joint survival of hemiarthroplasty versus total shoulder arthroplasty in osteoarthritis and rheumatoid arthritis. *Bone Jt J* 2019; **101 B**: 454–60.
- 8 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151**: 264–9.
- 9 Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A. How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry

381 reports with more than 15 years of follow-up. *Lancet* 2019; **393**: 647–54.

382 10 Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee
383 replacement last? A systematic review and meta-analysis of case series and national registry
384 reports with more than 15 years of follow-up. *Lancet* 2019; **393**: 655–63.

385 11 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for
386 systematic reviews. *Syst Rev* 2016; **5**: 210.

387 12 Gooberman-Hill R, Burston A, Clark E, *et al.* Involving patients in research: considering good
388 practice. *Musculoskeletal Care* 2013; **11**: 187–90.

389 13 Cohen J. Statistical power analysis for the behavioral sciences. Routledge, 2013.

390 14 Wylde V, Beswick AD, Dennis J, Gooberman-Hill R. Post-operative patient-related risk factors
391 for chronic pain after total knee replacement: a systematic review. *BMJ Open* 2017; **7**: e018105.

392 15 Khan A, Bunker TD, Kitson JB. Clinical and radiological follow-up of the Aequalis third-
393 generation cemented total shoulder replacement: A minimum ten-year study. *J Bone Jt Surg - Ser*
394 *B*; **91**: 1594–600.

395 16 Levy O, Tsvieli O, Merchant J, *et al.* Surface replacement arthroplasty for glenohumeral
396 arthropathy in patients aged younger than fiftyyears: Results after a minimum ten-year follow-up.
397 *J Shoulder Elb Surg* 2015; **24**: 1049–60.

398 17 Raiss P, Schmitt M, Bruckner T, *et al.* Results of cemented total shoulder replacement with a
399 minimum follow-up of ten years. *J Bone Jt Surg - Ser A*; **94**: e171.1-e171.10.

400 18 Sperling JW, Cofield RH, Rowland CM. Neer hemiarthroplasty and Neer total shoulder
401 arthroplasty in patients fifty years old or less: Long-term results. *J Bone Jt Surg - Ser A* 1998; **80**:
402 464–73.

- 403 19 Tammachote N, Sperling JW, Vathana T, Cofield RH, Harmsen WS, Schleck CD. Long-term
404 results of cemented metal-backed glenoid components for osteoarthritis of the shoulder. *J Bone Jt*
405 *Surg - Am Vol*; **91**: 160–6.
- 406 20 Young A, Walch G, Boileau P, *et al.* A multicentre study of the long-term results of using a flat-
407 back polyethylene glenoid component in shoulder replacement for primary osteoarthritis. *J Bone Jt*
408 *Surg - Ser B* 2011; **93 B**: 210–6.
- 409 21 Schoch B, Schleck C, Cofield RH, Sperling JW. Shoulder arthroplasty in patients younger than
410 50years: Minimum 20-year follow-up. *J Shoulder Elb Surg*; **24**: 705–10.
- 411 22 Rai P, Davies O, Wand J, Bigsby E. Long-term follow-up of the Copeland mark III shoulder
412 resurfacing hemi-arthroplasty. *J Orthop* 2016; **13**: 52–6.
- 413 23 Somerson JS, Matsen FA. Functional outcomes of the ream-And-run shoulder arthroplasty : A
414 concise follow-up of a previous report. *J Bone Jt Surg - Am Vol* 2017; **99**: 1999–2003.
- 415 24 Australian Orthopaedic Association National Joint Replacement Registry. Hip, Knee & Shoulder
416 Arthroplasty Annual Report 2019. 2019. <https://aoanjrr.sahmri.com/annual-reports-2019>.
- 417 25 Gauci MO, Bonneville N, Moineau G, Baba M, Walch G, Boileau P. Anatomical total shoulder
418 arthroplasty in young patients with osteoarthritis: all-polyethylene versus metal-backed glenoid.
419 *Bone Joint J*; **100**: 485–92.
- 420 26 Gerber C, Canonica S, Catanzaro S, Ernstbrunner L. Longitudinal observational study of reverse
421 total shoulder arthroplasty for irreparable rotator cuff dysfunction: results after 15 years. *J*
422 *Shoulder Elb Surg*; **27**: 831–8.
- 423 27 Neer 2nd CS, Watson KC, Stanton FJ. Recent experience in total shoulder replacement. *JBJS*
424 1982; **64**: 319–37.
- 425 28 Craig RS, Lane JCE, Carr AJ, Furniss D, Collins GS, Rees JL. Serious adverse events and lifetime

426 risk of reoperation after elective shoulder replacement: population based cohort study using
 427 hospital episode statistics for England. *BMJ* 2019; **364**: 1298.

428 29 Singh JA, Sperling JW, Cofield RH. Revision surgery following total shoulder arthroplasty:
 429 Analysis of 2588 shoulders over three decades (1976 to 2008). *J Bone Jt Surg - Ser B* 2011; **93 B**:
 430 1513–7.

431 30 Singh JA, Sperling JW, Cofield RH. Risk factors for revision surgery after humeral head
 432 replacement: 1,431 shoulders over 3 decades. *J Shoulder Elb Surg*; **21**: 1039–44.

433 31 National Joint Registry for England, Wales, Northern Ireland, the Isle of Man and the States of
 434 Guernsey. Prosthesis used in hip, knee, ankle, elbow and shoulder replacement procedures 2018.
 435 15th Annual Report 2019. 2019. <https://reports.njrcentre.org.uk/downloads>.

436 32 Norwegian National Advisory Unit on Arthroplasty and Hip Fractures Norwegian Report June
 437 2018. 2018. http://nrlweb.ihelse.net/eng/Rapporter/Report2018_english.pdf.

438 33 New Zealand Orthopaedic Association (NZOA) Joint Registry 20 year Annual Report. 2019.
 439 https://nzoa.org.nz/system/files/DH8328_NZJR_2019_Report_v4_7Nov19.pdf.

440 34 Werthel JD, Schoch B, Adams JE, Schleck C, Cofield R, Steinmann SP. Hemiarthroplasty Is an
 441 Option for Patients Older Than 70 Years With Glenohumeral Osteoarthritis. *Orthopedics*; **41**:
 442 222–8.

443 35 Sayers A, Evans JT, Whitehouse MR, Blom AW. Are competing risks models appropriate to
 444 describe implant failure? *Acta Orthop* 2018; **89**: 256–8.

445 36 Ollivere B, Metcalfe D, Perry DC, Haddad FS. SEARCHeD: Supporting Evaluation, Analysis and
 446 Reporting of routinely Collected Healthcare Data. *Bone Joint J.* 2020; **102-B**: 145–7.

447 37 Simovitch R, Flurin P-H, Wright T, Zuckerman JD, Roche CP. Quantifying success after total
 448 shoulder arthroplasty: the minimal clinically important difference. *J shoulder Elb Surg* 2018; **27**:

449 298–305.

450 38 Williamson PR, Altman DG, Bagley H, *et al.* The COMET Handbook: Version 1.0. *Trials*. 2017;
451 **18**: 1–50.

452 39 International Consortium for Health Outcome Measurement (ICHOM). <https://www.ichom.org/>
453 (accessed March 12, 2020).

454 40 Ramiro S, Page MJ, Whittle SL, *et al.* The OMERACT core domain set for clinical trials of
455 shoulder disorders. *J Rheumatol* 2019; **46**: 969–75.

456

Style Definition: Heading 2: Line spacing: 1.5 lines

Style Definition: Heading 3: Line spacing: 1.5 lines

How long does a shoulder replacement last? A systematic review and meta-analysis of ~~case-series~~case-series and national registry reports with more than 10 years of follow-up

Jonathan P Evans ^{MD^{1,2}}, Jonathan T Evans ^{MD³}, Richard S Craig ^{FRCS (T&O)^{4,5}}, Hassan R Mohammad ^{MRes^{4,5}}, Adrian Sayers ^{MSc³}, Prof Ashley W Blom ^{PhD^{3,6}}, Michael R Whitehouse ^{PhD^{3,6}}, Prof Jonathan L Rees ^{MD^{4,5}}

1. Health & Policy Research Group, University of Exeter, Exeter, EX1 2LU, UK
2. National Institute for Health Research Applied Research Collaboration South West Peninsula, University of Exeter, Exeter, EX1 2LU, UK
3. Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, 1st Floor Learning & Research Building, Southmead Hospital, Bristol, BS10 5NB, UK
4. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, OX3 7LD, UK
5. National Institute for Health Research Oxford Biomedical Research Centre, Oxford, OX3 7LD, UK
6. National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, BS10 5NB, UK

Correspondence to:

Mr Jonathan Peter Evans

Health & Policy Research Group, University of Exeter, Exeter, EX1 2LU, UK

j.p.evans2@exeter.ac.uk

31 Panel: Research in context

32

33 **Evidence before this study**

34 Survival of shoulder replacements has often been reported in small ~~case-series~~[case-series](#), with some follow-
35 up extending beyond 20 years, however individual case-series are prone to bias and reporting has been
36 highly heterogeneous. We searched MEDLINE and Embase for systematic reviews and meta-analyses of
37 shoulder replacement series that were published in English. Of the 37 systematic reviews we identified, no
38 articles reported combined survival estimates or patient reported outcome measures with more than 10 years
39 follow-up. A previous analysis of the UK Hospital Episode Statistics (HES) dataset, published in 2019,
40 combined all types of shoulder implants and found overall survival to be 90·0% (95% CI 89·6% to 90·3%)
41 at 10 years. No study to date has attempted to provide pooled survival estimates and pooled patient reported
42 outcomes for shoulder replacements more than 10 years after surgery.

43 **Added value of this study**

44 To our knowledge, we provide the first pooled survival estimate, drawn from multiple sources, for shoulder
45 replacements at 10 years. We have also shown that shoulder replacements have a sustained positive impact
46 on patients' lives to 10 years after surgery. Our findings showed that approximately 92% of total shoulder
47 replacements, 91% of shoulder humeral hemiarthroplasties and 94% of reverse total shoulder replacements
48 last for 10 years.

49 **Implications of all the available evidence**

50 Our findings provide valuable and overdue information for patients and clinicians considering shoulder
51 replacement surgery. It is the first study to provide a simple and generalizable answer to two very important
52 questions: "How long does a shoulder replacement last?" and "Will my shoulder be better in the long-term
53 after surgery?" The data will also be useful for those commissioning healthcare services.

54

55 Abstract

56 Background

57 Shoulder replacement is an increasingly common treatment for end-stage degenerative shoulder conditions.
58 ~~Given time, all~~ Some shoulder replacements will fail ~~and further operations may be required~~. It is important
59 for patients and clinicians to know how long shoulder replacements last and how effectively they improve
60 pain and function. This study aims to determine the longevity and long-term efficacy of shoulder
61 replacements.

62 Methods

63 In this systematic review and meta-analysis, we searched MEDLINE and Embase for articles reporting 10-
64 year or greater survival of Total Shoulder Replacements (TSR), Humeral Hemiarthroplasties (HA) and
65 Reverse Total Shoulder Replacements (RTSR). Survival, implant and Patient Reported Outcome Measures
66 (PROMs) data were extracted. National joint replacement registries were reviewed and analysed separately.
67 We weighted each series and calculated a pooled survival estimate at 10, 15 and 20 years. For PROMs we
68 pooled the Standardised Mean Difference (SMD) at 10 years.

69 Findings

70 We identified 10 series reporting all-cause survival of 529 TSRs and 420 HA, no series for RTSR met our
71 inclusion criteria. The estimated 10-year survival for TSR was 95·6% (95% CI 93·6, 97·6) and HA 90·4%
72 (95% CI 87·0, 94·0). A single registry contributed 7941 TSRs, 3495 HAs and 8049 RTSRs. The pooled
73 registry 10-year survival for TSR was 92·0% (95% CI 91·0, 93·0), HA 90·5% (95% CI 81·8, 95·1) and
74 RTSR 94·4% (95% CI 93·1, 95·7) for osteoarthritis and 93·6% (95% CI 91·0, 95·4) for rotator cuff
75 arthropathy. Pooled 10-year PROMs revealed a substantial improvement from baseline scores (SMD 2·13
76 95% CI 1·93, 2·34).

77 Interpretation

78 Over 90% of shoulder replacements last more than 10 years and patient reported benefits are sustained.
79 This long overdue information will be of use to patients and health-care providers.

80 Funding

81 The National Institute for Health Research, the National Joint Registry for England, Wales, Northern
82 Ireland, and Isle of Man, and the Royal College of Surgeons of England.

83 Introduction

84 Patients with severe pain and disability from degenerative shoulder conditions want to know whether they
85 will benefit from shoulder replacement surgery, which type of replacement may be best and what they can
86 expect in the long-term following surgery.¹ [A review of seven national arthroplasty registers in 2017](#)
87 [suggested there has been](#) a secular increase in the number of shoulder replacements performed for
88 patients with [both](#) osteoarthritis and rotator cuff tear ~~arthropathy~~ [arthropathy](#). [Overall the annual](#)
89 [incidence rate has increased 2.8 timesfold in the last decade, but significant variation exists between](#)
90 [countries](#).² There is a paucity of high quality outcome data to aid joint decision making by patients and
91 clinicians, and to assist both commissioners and providers in understanding the utility and likely revision
92 burden associated with undertaking these procedures.

93 Available randomised controlled trials (RCTs) are particularly limited, by size and design, in their ability
94 to evaluate the longer-term outcomes and risks of primary shoulder arthroplasty, in particular the
95 requirement for revision surgery.³ To better understand the long-term benefits and risks of shoulder
96 replacement surgery for these patients, a formal appraisal and synthesis of the more frequently available
97 non-randomised study data is needed.

98 Ideally, clinicians and surgeons should be able to provide patients with contemporary condition-, age- and
99 implant-specific outcome data for any proposed procedure and available alternatives. While implant
100 manufacturers do facilitate the collection of implant-level data in order to gain relevant benchmark
101 accreditation,⁴ detailed and reliable data are not yet available for shoulders. Until such granular brand-level
102 information is available, clinicians and patients need accurate information on classes of available implants.
103 Hip and knee replacement have shown that although there is variation between brands, classes of implants
104 behave in broadly similar fashion.^{5,6} The three main constructs or classes available and referred to in this
105 study are conventional total shoulder replacement (TSR), humeral hemiarthroplasty (HA), and reverse total
106 shoulder replacement (RTSR). There is likely to be heterogeneity between indications for surgery,
107 mechanisms of failure and overall revision rates between these different constructs.⁷

108 In this study we sought to answer a simple but important question posed by all patients: How long does a
109 shoulder replacement last? ~~We planned to make the~~ We aimed to provide the best quality pooled estimates
110 of implant survival at a minimum 10 years' follow-up. The decision to revise a poorly performing shoulder
111 replacement is multifactorial that may be sensitive to both patient and surgeon preferences. Therefore, we
112 also ~~aimed~~~~sought~~ to make a pooled estimate of the likely patient reported outcome at long-term follow-up,
113 in essence to answer the question: Will my shoulder be better 10 years after surgery?

114

115 Methods

116 Search strategy and selection criteria

117 We conducted a systematic review and meta-analysis assessing the survival of shoulder replacements in
118 case-series and national joint registries following a predefined protocol registered with PROSPERO
119 ([CRD42019140221](#)~~CRD4201910221~~) and complying with PRISMA guidelines.⁸

120 A search strategy using keywords and MeSH terms relating to shoulder replacement and survival (appendix
121 1) was used in the databases MEDLINE and Embase accessed through OVID Silver Platter. The databases
122 were searched from their commencement to 24th September 2019. The strategy development was guided
123 by previously published search strategies exploring the survival of hip and knee replacements.^{9,10} Manual
124 screening of the bibliographies of the full-text articles and systematic reviews was also undertaken.

125 Studies were included if they assessed patients who had undergone any type of shoulder replacement (a
126 total shoulder replacement (TSR), humeral hemiarthroplasty (HA) or reverse total shoulder replacement
127 (RTSR)). Humeral components (stemmed, stemless or resurfacing) were all considered as TSR or HA
128 dependent on whether the glenoid (shoulder socket) was replaced or not and not sub-classified. The
129 indication (reason) for surgery had to be predominantly osteoarthritis (OA) or rotator cuff arthropathy
130 (RTCA). For inclusion, the case-series or published registry report had to report the survival of a specific
131 brand of implant with a mean or median follow-up of greater than 10 years. It is widely accepted that
132 survival of hip arthroplasties varies by the brand of implant.⁵ Although this has not specifically been
133 assessed in shoulder replacements, the technique of treating each brand as its own series ~~is methodology~~
134 was utilised, as the assumption that variation in survival by brand that exists in hip and knee replacements,
135 therefore the assumption would seem sensible for shoulder replacements as well. ~~W~~weighting of implants
136 in the meta-analysis would therefore ~~provide~~s the most robust survival estimates. This allows us to treat
137 each series as an individual study and weight the meta-analysis of survival results according to the standard
138 error of each series. Aggregate data from multiple implant brands would not allow this granularity and thus
139 hide the potential variability in performance between implant brands. A cut-off of minimum mean or

median follow-up of 10 years was chosen as the subject of interest of this study was “long-term” survival, where there is a current paucity of information. We accept this definition may vary subjectively but 10 years allowed inclusion of sufficient studies to make analyses robust and represents a time period that is relatable to patients and clinicians.

Studies were excluded if they reported the outcome of revision surgery, as this is often more complex surgery and carries different survivorship. Conference abstracts were excluded due to the limited data available from these reports. Systematic reviews were assessed for their citations but did not include their pooled data to avoid duplication.

The reports from all available national joint registers that collect and publish the individual implant-specific survivorship for shoulder replacements with at least 10-years of follow-up were assessed. Reports were identified through the systematic search if published or accessed through their websites, ~~or published reports.~~

Article screening and data extraction

Screening was undertaken in a stepwise manner using the web application Rayyan.¹¹ Journal article titles and abstracts were screened by two reviewers (JTE and HM) with arbitration of conflict undertaken by JPE. Full-text review and data extraction were undertaken by two reviewers independently (JPE and JTE). Data extracted were: publication date, baseline population demographics, number of patients (n), surgical indication proportion (% OA and/or % RCTA), follow-up duration (>10 years), implant name and construct type (TSA, HA or RTSA), loss to follow-up, survival estimates (including CIs) and all available Patient Reported Outcome Measure (PROM) (e.g. Visual Analogue Scales (VAS), Constant score, Disabilities of the Arm, Shoulder and Hand (DASH)), data ~~(outcome measure used)~~ (outcome measure used) baseline mean score (SD), follow-up duration in 5 year increments, follow-up mean score (SD)). Data were not extracted from figures (e.g. Kaplan Meier plots) to avoid potential transcription inaccuracy. Discrepancy in extracted data was discussed by the authors, following which there were no cases of disagreement.

164 Statistical Analysis

165 For the assessment of the published case-series our primary exposure was the shoulder replacement implant
166 and our primary outcome was all-cause revision, of any part of this construct, as guided by our patient
167 group.¹² Statistical analysis was performed with Stata 15 (*Stata Statistical Software: Release 15*. College
168 Station, TX: StataCorp LLC). Survival estimates, assuming that survivorship approximated revision risk,
169 were pooled by meta-analysis. Each series was weighted according to its standard error (calculated from
170 published confidence intervals). The effect size (Standardised Mean Difference (SMD)) of the primary
171 PROMs reported in each study was pooled with meta-analysis with weighting according to sample size and
172 analysed using a random effects model as a more conservative estimate of treatment effect. Effect size was
173 considered small if it was less than ≥ 0.2 , moderate if ≥ 0.5 and large if ≥ 0.8 .¹³

174 Quality assessment

175 Study quality was assessed using the non-summative four-point system (consecutive cases, multi-centre,
176 under 20% loss to follow-up and use of multivariable analysis) developed by Wylde et al.¹⁴ This was
177 selected in preference to the summative MINORS score due to the high loss to follow-up in joint
178 replacement case-series and because some of the scoring criteria in MINORS were not relevant to joint
179 replacement.

180

181 Role of the funding source

182 The funder of the study had no role in study design, data collection, data analysis, data interpretation, or
183 writing of the report. All authors had access to the raw data. The corresponding author had full access to
184 all of the data and the final responsibility to submit for publication

Results

The search of published ~~case-series~~case-series produced 1,376 articles. Of these, 449 duplicates were removed, leaving 927 articles for screening (figure 1). After screening, 36 full-text articles were reviewed. Additional citation searches through previously published systematic review references yielded four further full-text reviews, none of which met the inclusion criteria. Following review of full-text articles, nine articles reporting 10 individual implant specific series were included in the survival analysis, six articles that reported both survival analysis and PROMs were included in the PROMs analysis. A summary of study level characteristics is provided in Table 1. The proportion of OA as the primary surgical indication was 59% for TSR and 48% for HA. The reporting of indication was variable and was interpretable in only seven articles.

Quality assessment revealed that six (60%) of the 10 series were consecutive, two (20%) were multicenter, nine (90%) had >80% follow-up (at an average with mean loss to follow up of 8.4%, ranging from 0% to 23.7%), and none undertook multivariable analysis. These proportions are in keeping with the fact that the quality of published case-series is low.

Case-series

Six unique series, published between 1998 – 2015, reported survival of 529 total shoulder replacements (TSR) at 13 time points with follow-up ranging from 10 to 21 years (Appendix 2).¹⁵⁻²¹ Four reported survival at exactly 10 years (466 TSRs), three reported survival at 15 years (427 TSRs) and one reported survival at 20 years (19 TSRs). Pooled survival from those studies reporting at exactly 10 years was 95.6% (95% CI 93.6, 97.6) at 15 years 88.5% (95% CI 83.4, 94.1) and at 20 years 83.2% (95% CI 70.5, 97.8) (figure 2). When studies reported survival estimates at between >10 and ≤15 years, these results were rounded down to 10 years as a sensitivity analysis. This ~~which~~ did not report outcomes at exactly 10 years, rounding down of reported survival between >10 to <15 years from those series resulted in a pooled survival of six series (529 TSRs) of 90.0% (95% CI 88.3, 91.7) (figure 3).

Four unique series, published between 1998 – 2017, reported survival of 364 shoulder humeral hemiarthroplasties (HAs) at 10 time points with follow-up ranging from 10 to 21 years (Appendix 2).^{16,18,21–23} Three reported survival at exactly 10 years (327 HAs), two at 15 years (151 HAs) and one at 20 years (56 HAs). Pooled survival at exactly 10 years was 90.4% (95% CI 87.0, 94.0), at 15 years 90.6% (95% CI 84.1, 97.1), and at 20 years 75.6% (95% CI 65.9, 86.5) (figure 2). Rounding down of reported survival from those series closest to ≥10 but <15 years resulted in a pooled survival of four series (364 HAs) of 92.5% (95% CI 89.6, 95.3) (figure 3).

No unique single implant series with a mean follow-up of at least 10 years were found for reverse total shoulder replacements (RTSA).

Registry data

The reports of implant-level data at 10 years were only available from a single registry, the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) 2019 annual report.²⁴ This report yielded 10-year survival of eight series of TSRs (7,941 arthroplasties), eight series of HAs (3,495 arthroplasties) and five series of RTSRs (8,049 arthroplasties). Pooled survival estimates from registry data for TSRs at 10 years were 92.0% (95% CI 91.0, 93.0); for HAs 90.5% (95% CI 81.8, 95.1) and for RTSR were 94.4% (95% CI 93.1, 95.7) for a primary diagnosis of OA, and 93.6% (95% CI 91.0, 95.4) for a diagnosis of RTCA (single implant reported) (figure 4).

Patient Reported Outcome Measures

Of the 14 studies reporting survival analysis, six reported the implant level PROMs of 617 shoulder replacements for inclusion in the PROMs meta-analysis; this included two studies not included in the survival meta-analysis, excluded as they did not report confidence intervals.^{17,19,20,23,25,26} Four studies reported PROMs on TSR, one on RTSR and one on HA. All reported the outcome of shoulder-specific PROMs, without the addition of generic quality of life measures. Five studies reported the Constant score, one the simple shoulder test (SST) and one a four-point linear pain scale previously described by Neer.²⁷ Pooled PROMs data showed a large effect of improved outcome from baseline (SMD 2.13 95% CI 1.93,

234 2·34) (figure 5). Subgroup analysis of PROMs exclusively from TSRs reduced the effect size marginally
235 (SMD 2·02 95% CI 1·86, 2·19). Implant-level 10-year PROMs were not published in any registry reports.
236 The New Zealand registry report 10-year PROMs, which were categorised by construct only (TSR, HA,
237 RTSR, Partial resurfacing of head). Although no baseline PROMs are available for comparison, at 10-years
238 the Oxford Shoulder Score (OSS) mean for all implants was 39.1/48 (95% CI 38.4, 39.8), for TSA (n=335)
239 41.0/48 (40.0, 42.0), HA (n=104) 39.4/48 (37.7, 41.1), RTSR (n=104) 39.4 (37.7, 41.1).

Discussion

We found that 90% of shoulder replacements last for at least 10 years and that patients can expect a large and sustained improvement in their patient reported outcome measures.

The methodology used is one that has been previously applied successfully to hip and knee replacement,^{9,10} with the production of simple and generalisable results. The application of this process to shoulder replacement proved more complex due to sparsity and heterogeneity of data and highlights why the study question has not previously been answered. However, despite these limitations, the data from both registries and case-series independently estimate the same results. This is encouraging and suggests that these ~~case series~~ case-series are not subject to selection and publication bias.

The methods applied in this study use an individual estimate for each implant series, which is then synthesised to provide single pooled construct estimate weighted according to the standard error. ~~This analysis, whereby an estimate is produced depended on how frequently an implant is used, has not previously been undertaken in the analysis of shoulder replacements. The implant has been shown to be fundamental to the survival outcome of hip and knee replacement and is likely to be just as important in shoulder replacement and each individual series should be considered as a different patient cohort.⁵ We have used the individual estimates for each implant to synthesise a single pooled estimate, weighting the estimates according to standard error. This type of analysis, deriving an overall estimate according to how frequently each implant has been used, is unique to our study.~~ This analysis is dependent upon case-series, and registries' reporting of implant level data, as the only method where the patterns of implant failure can be accounted for, ~~under the assumption that different types of implant demonstrate different survival estimates.⁵~~

Implant survival at more than 10 years was greater than 90% for both TSR and HA in the case-series data, and also in the Australian registry data. This finding is concordant with the limited number of extended survival reports using multi-implant cohorts, including the assessment of Hospital Episode Statistics (HES) data in England²⁸ of 90% (95%CI 89.6 - 90.3) in a combined arthroplasty cohort, and Mayo clinic registry

265 data^{29,30} of 90.2% (95% CI 88.7, 91.7) for TSR and 90.0% (95% CI 88.0, 92.0) for HA. This study found
266 very limited extended case-series 20-year data, all from the Mayo group, with survival for TSRs of 83.2%
267 and HAs 75.6%, which are lower than the HES report of 87.8% (95% CI 87.2, 88.4) at 18 years but
268 comparable to the full Mayo Clinic registry of 81.4% (95% CI 78.4, 84.5) for TSR, but worse than the HA
269 survival of 85.0% (95% CI 81.8, 88.4) at a 20 years, notably there is a younger age cohort in their HA
270 case-series. It is notable that the demographic characteristics from the case-studyseries and registry data are
271 similar for the TSR group, and concordantly their survival rates are also comparable. For the HA group, the
272 case-studyseries data contain a more male dominated and younger population. All but one of the case-
273 studiesseries report an average age of <60yrs, therefore the survival findings from case-series may lack
274 generalisability.

275 For RTSR, there was an absence of any implant level data from ~~case-series~~case-series at more than 10 years.
276 This is concerning as it is currently utilized in over 50% of shoulder replacements in the UK, Norway,
277 Australia and New Zealand.^{24,31-33} It is surprising that this change in practice has occurred so rapidly with
278 such paucity of long-term outcome evidence, particularly after the well documented problems with the
279 widespread adoption of unproven technology in joint replacement.⁶ It is therefore reassuring that we have
280 been able to assess survival of RTSR at 10 years using data synthesised from the Australian registry data
281 which reveals a survival of 94.0% (95% CI 93.1, 95.7) for OA and 93.6% (95% CI 91.0, 95.4) for RTCA.

282 Of the studies that reported survival of shoulder replacements at a mean of >10 years, five did not include
283 confidence intervals and could not be added to the meta-analysis, six reported the composite survival of
284 cohorts that included multiple different implants. Addition of these data would have resulted in the inclusion
285 of 1,482 arthroplasties, increasing the analysis cohort by >150%. Failure of individual components of the
286 construct (e.g. the glenoid or humeral component in isolation) was also reported in a large series that was
287 excluded from the meta-analysis owing to the absence of an all-cause construct survival estimate.³⁴
288 Although component-failure data are of interest, we would regard this as best reported as a secondary
289 endpoint, with the all-cause 1-Kaplan Meier estimate as the most appropriate method of reporting

290 survivorship, which should always include the number of shoulder replacements at risk at the time of
291 reporting.³⁵

292 As shoulder replacement registries may not provide long-term survival for some time to come, we remain
293 somewhat reliant on case-series data. If these series are to reliably inform the surgical community of
294 implants at risk, they must be transparently reported according to current guidance on the reporting of
295 healthcare data.³⁶ As novel implants and techniques are developed, we will also continue to be reliant upon
296 case-series to highlight potential improvements in survivorship and function.

297 This study has identified that at over >10 years from the primary intervention a large improvement (SMD
298 2·13) in PROMs scores was maintained. A linear transformation, making all scores interpretable from the
299 Constant score scale, also demonstrates a mean change score of 40·4, which exceeds the minimal clinically
300 reported difference (MCID) of 12·8 ± 2·5 points for TSR.³⁷ The authors recognise the concern regarding
301 the validity of the Constant score, and suggest that future studies report PROMs with proven validity and
302 responsiveness. The New Zealand registry provided the only published comparator of construct-level, but
303 not implant level, PROMs data. At 10 years this was limited to 674 replacements. Their high OSS at 10
304 years (80% of total score) does suggest a sustained benefit of shoulder replacements. As the New Zealand
305 registry does not provide baseline pre-operative scores, comparison of SMD could not be undertaken.

306 We echo the calls for consensus in outcome choice to facilitate synthesis of data. Initiatives that promote
307 the use of core outcome sets include the Core Outcome Measures in Effectiveness Trials (COMET),
308 Outcome Measures in Rheumatology (OMERACT) and the International Consortium for Health Outcome
309 Measurement (ICHOM).^{38–40} Furthermore, the inclusion of PROMs in registry data has the potential to
310 dramatically improve the assessment of patient-focused outcomes. Currently, clear associations between
311 survival of a shoulder implant and the patient-focused domains of pain, function and quality of life cannot
312 be ascertained.

313 There are limitations of this work. The data did not allow stratification or adjustment for patient factors that
314 may have affected outcomes in the pooled analysis. The analysis could not account for differing thresholds
315 for revision between surgeons. It is notable that many of the historic series are derived from single-surgeon
316 series and therefore surgeon preferences may alter the resultant weighted synthesis of survivorship. We also
317 recognise that emergent techniques and implants may demonstrate superior (or inferior) survivorship and
318 function that is yet to be demonstrated with long-term follow-up. The impact of historic series that have
319 utilised implants subsequently recognised as having worse outcomes can affect a synthesis of long-term
320 outcomes. The series from Levy et al ¹⁶ which included metal-backed glenoid components had a large
321 weighting that reduced the overall survival estimate. Reporting early failure of certain implants is important
322 and for the best available overall estimates should continue to be included. As not all failure results in
323 revision, we reported patient-reported outcomes to better define the overall value of shoulder replacement.
324 Our pooled registry results are drawn exclusively from the Australian register. As the available follow-up
325 in other registries increases, a wealth of data will soon become available, and we would encourage implant
326 level reporting by brand and product line. We also assumed that survival estimates are equivalent to risks
327 for generating pooled estimates, and although the assumption that no censoring occurs (patients dying with
328 a shoulder in situ) is violated, it provides a useful method of aggregation in the absence of individual patient
329 data. The aggregated estimates of survival are however the largest possible sample and this is the largest
330 report of this type and length of follow-up.

331 The strengths of this study include an inclusive and comprehensive design and realistic interpretation of
332 survivorship that accounts for all revisions and not a limited or biased subset, as well as a patient outcome
333 focus. From a patient perspective, all revision surgery carries risk and therefore all-cause revision should
334 be considered.

335 Conclusion

336 By pooling survival from case-series and registry data, we have been able to provide a reliable estimate of
337 10-year survival rate of shoulder replacements. We found that over 90% of shoulder replacements last for

338 at least 10 years. Patients experienced sustained and marked benefit to 10 years. [This information should](#)
339 [be reassuring for patients, health professionals and commissioners of health services.](#)

340

341 **Contributions**

342 JPE and JTE were responsible for study concept, design, screening, data extraction, data analysis, and
343 writing of this manuscript.

344 HM completed the primary screening of abstracts and review of the manuscript.

345 JR, AB, MRW, RC and AS were responsible for study concept, design, and writing of the manuscript.

346

347 **Declaration of interests**

348 We declare no competing interests.

349

350 **Acknowledgments**

351 This report is independent research supported by the National Institute for Health Research Applied
352 Research Collaboration South West Peninsula, National Institute for Health Research Biomedical
353 Research Centre at the University Hospitals Bristol National Health Service (NHS) Foundation Trust and
354 the University of Bristol, and National Institute for Health Research Oxford Biomedical Research Centre.
355 The views expressed in this publication are those of the author(s) and not necessarily those of the
356 National Institute for Health Research or the Department of Health and Social Care.

357

358 **Funding**

359 This study was supported by the NIHR Applied Research Collaboration South West Peninsula, NIHR
360 Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and University of
361 Bristol, and NIHR Oxford Biomedical Research Centre.

362 JPE was supported by an NIHR Clinical Lectureship.

363 JTE was supported by the joint National Joint Registry of England, Wales, Northern Ireland and the Isle
364 of Man and Royal College of Surgeons of England Fellowship.

365 AS was supported by an MRC Strategic Skills Fellowship MR/L01226X/1

366

367 References

- 368 1 Rangan A, Upadhaya S, Regan S, Toye F, Rees JL. Research priorities for shoulder surgery:
369 Results of the 2015 James Lind Alliance patient and clinician priority setting partnership. *BMJ*
370 *Open* 2016; **6**: e010412.
- 371 2 Lübbecke A, Rees JL, Barea C, Combescure C, Carr AJ, Silman AJ. International variation in
372 shoulder arthroplasty. *Acta Orthop* 2017; **88**: 592–9.
- 373 3 Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL. Shoulder replacement surgery for
374 osteoarthritis and rotator cuff tear arthropathy. *Cochrane Database Syst Rev* 2020.
- 375 4 Orthopaedic Data Evaluation Panel (ODEP). <http://www.odep.org.uk/> (accessed March 11, 2020).
- 376 5 Deere KC, Whitehouse MR, Porter M, Blom AW, Sayers A. Assessing the non-inferiority of
377 prosthesis constructs used in hip replacement using data from the National Joint Registry of
378 England, Wales, Northern Ireland and the Isle of man: a benchmarking study. *BMJ Open* 2019; **9**:
379 e026685.
- 380 6 Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip
381 replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet*
382 2012; **379**: 1199–204.
- 383 7 Lapner PLC, Rollins MD, Netting C, Tuna M, Bader Eddeen A, van Walraven C. A population-
384 based comparison of joint survival of hemiarthroplasty versus total shoulder arthroplasty in
385 osteoarthritis and rheumatoid arthritis. *Bone Jt J* 2019; **101 B**: 454–60.
- 386 8 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and
387 meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151**: 264–9.
- 388 9 Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A. How long does a hip
389 replacement last? A systematic review and meta-analysis of case series and national registry

reports with more than 15 years of follow-up. *Lancet* 2019; **393**: 647–54.

10 Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet* 2019; **393**: 655–63.

11 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016; **5**: 210.

12 Gooberman-Hill R, Burston A, Clark E, *et al*. Involving patients in research: considering good practice. *Musculoskeletal Care* 2013; **11**: 187–90.

13 Cohen J. Statistical power analysis for the behavioral sciences. Routledge, 2013.

14 Wylde V, Beswick AD, Dennis J, Gooberman-Hill R. Post-operative patient-related risk factors for chronic pain after total knee replacement: a systematic review. *BMJ Open* 2017; **7**: e018105.

15 Khan A, Bunker TD, Kitson JB. Clinical and radiological follow-up of the Aequalis third-generation cemented total shoulder replacement: A minimum ten-year study. *J Bone Jt Surg - Ser B*; **91**: 1594–600.

16 Levy O, Tsvieli O, Merchant J, *et al*. Surface replacement arthroplasty for glenohumeral arthropathy in patients aged younger than fiftyyears: Results after a minimum ten-year follow-up. *J Shoulder Elb Surg* 2015; **24**: 1049–60.

17 Raiss P, Schmitt M, Bruckner T, *et al*. Results of cemented total shoulder replacement with a minimum follow-up of ten years. *J Bone Jt Surg - Ser A*; **94**: e171.1-e171.10.

18 Sperling JW, Cofield RH, Rowland CM. Neer hemiarthroplasty and Neer total shoulder arthroplasty in patients fifty years old or less: Long-term results. *J Bone Jt Surg - Ser A* 1998; **80**: 464–73.

- 412 19 Tammachote N, Sperling JW, Vathana T, Cofield RH, Harmsen WS, Schleck CD. Long-term
413 results of cemented metal-backed glenoid components for osteoarthritis of the shoulder. *J Bone Jt*
414 *Surg - Am Vol*; **91**: 160–6.
- 415 20 Young A, Walch G, Boileau P, *et al.* A multicentre study of the long-term results of using a flat-
416 back polyethylene glenoid component in shoulder replacement for primary osteoarthritis. *J Bone Jt*
417 *Surg - Ser B* 2011; **93 B**: 210–6.
- 418 21 Schoch B, Schleck C, Cofield RH, Sperling JW. Shoulder arthroplasty in patients younger than
419 50years: Minimum 20-year follow-up. *J Shoulder Elb Surg*; **24**: 705–10.
- 420 22 Rai P, Davies O, Wand J, Bigsby E. Long-term follow-up of the Copeland mark III shoulder
421 resurfacing hemi-arthroplasty. *J Orthop* 2016; **13**: 52–6.
- 422 23 Somerson JS, Matsen FA. Functional outcomes of the ream-And-run shoulder arthroplasty : A
423 concise follow-up of a previous report. *J Bone Jt Surg - Am Vol* 2017; **99**: 1999–2003.
- 424 24 Australian Orthopaedic Association National Joint Replacement Registry. Hip, Knee & Shoulder
425 Arthroplasty Annual Report 2019. 2019. <https://aoanjrr.sahmri.com/annual-reports-2019>.
- 426 25 Gauci MO, Bonneville N, Moineau G, Baba M, Walch G, Boileau P. Anatomical total shoulder
427 arthroplasty in young patients with osteoarthritis: all-polyethylene versus metal-backed glenoid.
428 *Bone Joint J*; **100**: 485–92.
- 429 26 Gerber C, Canonica S, Catanzaro S, Ernstbrunner L. Longitudinal observational study of reverse
430 total shoulder arthroplasty for irreparable rotator cuff dysfunction: results after 15 years. *J*
431 *Shoulder Elb Surg*; **27**: 831–8.
- 432 27 Neer 2nd CS, Watson KC, Stanton FJ. Recent experience in total shoulder replacement. *JBJS*
433 1982; **64**: 319–37.
- 434 28 Craig RS, Lane JCE, Carr AJ, Furniss D, Collins GS, Rees JL. Serious adverse events and lifetime

435 risk of reoperation after elective shoulder replacement: population based cohort study using
 436 hospital episode statistics for England. *BMJ* 2019; **364**: l298.

437 29 Singh JA, Sperling JW, Cofield RH. Revision surgery following total shoulder arthroplasty:
 438 Analysis of 2588 shoulders over three decades (1976 to 2008). *J Bone Jt Surg - Ser B* 2011; **93 B**:
 439 1513–7.

440 30 Singh JA, Sperling JW, Cofield RH. Risk factors for revision surgery after humeral head
 441 replacement: 1,431 shoulders over 3 decades. *J Shoulder Elb Surg*; **21**: 1039–44.

442 31 National Joint Registry for England, Wales, Northern Ireland, the Isle of Man and the States of
 443 Guernsey. Prosthesis used in hip, knee, ankle, elbow and shoulder replacement procedures 2018.
 444 15th Annual Report 2019. 2019. <https://reports.njrcentre.org.uk/downloads>.

445 32 Norwegian National Advisory Unit on Arthroplasty and Hip Fractures Norwegian Report June
 446 2018. 2018. http://nrlweb.ihelse.net/eng/Rapporter/Report2018_english.pdf.

447 33 New Zealand Orthopaedic Association (NZOA) Joint Registry 20 year Annual Report. 2019.
 448 https://nzoa.org.nz/system/files/DH8328_NZJR_2019_Report_v4_7Nov19.pdf.

449 34 Werthel JD, Schoch B, Adams JE, Schleck C, Cofield R, Steinmann SP. Hemiarthroplasty Is an
 450 Option for Patients Older Than 70 Years With Glenohumeral Osteoarthritis. *Orthopedics*; **41**:
 451 222–8.

452 35 Sayers A, Evans JT, Whitehouse MR, Blom AW. Are competing risks models appropriate to
 453 describe implant failure? *Acta Orthop* 2018; **89**: 256–8.

454 36 Ollivere B, Metcalfe D, Perry DC, Haddad FS. SEARCHeD: Supporting Evaluation, Analysis and
 455 Reporting of routinely Collected Healthcare Data. *Bone Joint J.* 2020; **102-B**: 145–7.

456 37 Simovitch R, Flurin P-H, Wright T, Zuckerman JD, Roche CP. Quantifying success after total
 457 shoulder arthroplasty: the minimal clinically important difference. *J shoulder Elb Surg* 2018; **27**:

458 298–305.

459 38 Williamson PR, Altman DG, Bagley H, *et al.* The COMET Handbook: Version 1.0. *Trials*. 2017;
460 **18**: 1–50.

461 39 International Consortium for Health Outcome Measurement (ICHOM). <https://www.ichom.org/>
462 (accessed March 12, 2020).

463 40 Ramiro S, Page MJ, Whittle SL, *et al.* The OMERACT core domain set for clinical trials of
464 shoulder disorders. *J Rheumatol* 2019; **46**: 969–75.

465

Table 1: Study-level and participant-level characteristics of contributing data sources * estimates from whole group data.
[‡]weighted age and %female dependent on number of arthroplasties in the series. TSR – Total Shoulder Replacement, HA – Hemiarthroplasty, RTSR – Reverse Total Shoulder Replacement

Individual case series articles				Australian Orthopaedic Association National Joint Replacement Registry annual report 2019		
	TSR	HA	RTSR	TSR	HA	RTSR
Study level characteristics						
Location	UK (2), Germany (1), USA (2), Pan European (1)	UK (2) USA (3)	0	Australia	Australia	Australia
Number of unique implant series included	6	4	0	8	8	5
Year of publication	1998-2015	1998 - 2017	NA	2019	2019	2019
Participant level characteristics						
Mean age (years)	66.9 [‡]	54.5 [‡]	NA	72.2*	67.7*	74.3*
% of female patients	57.8 [‡]	34.4 [‡]	NA	61.6*	63.4*	64.3*
Total Arthroplasties (n) at start	529	420		7941	3495	8049
Loss to follow-up	62 (11.7%)	30 (7.1%)				

Figure 1

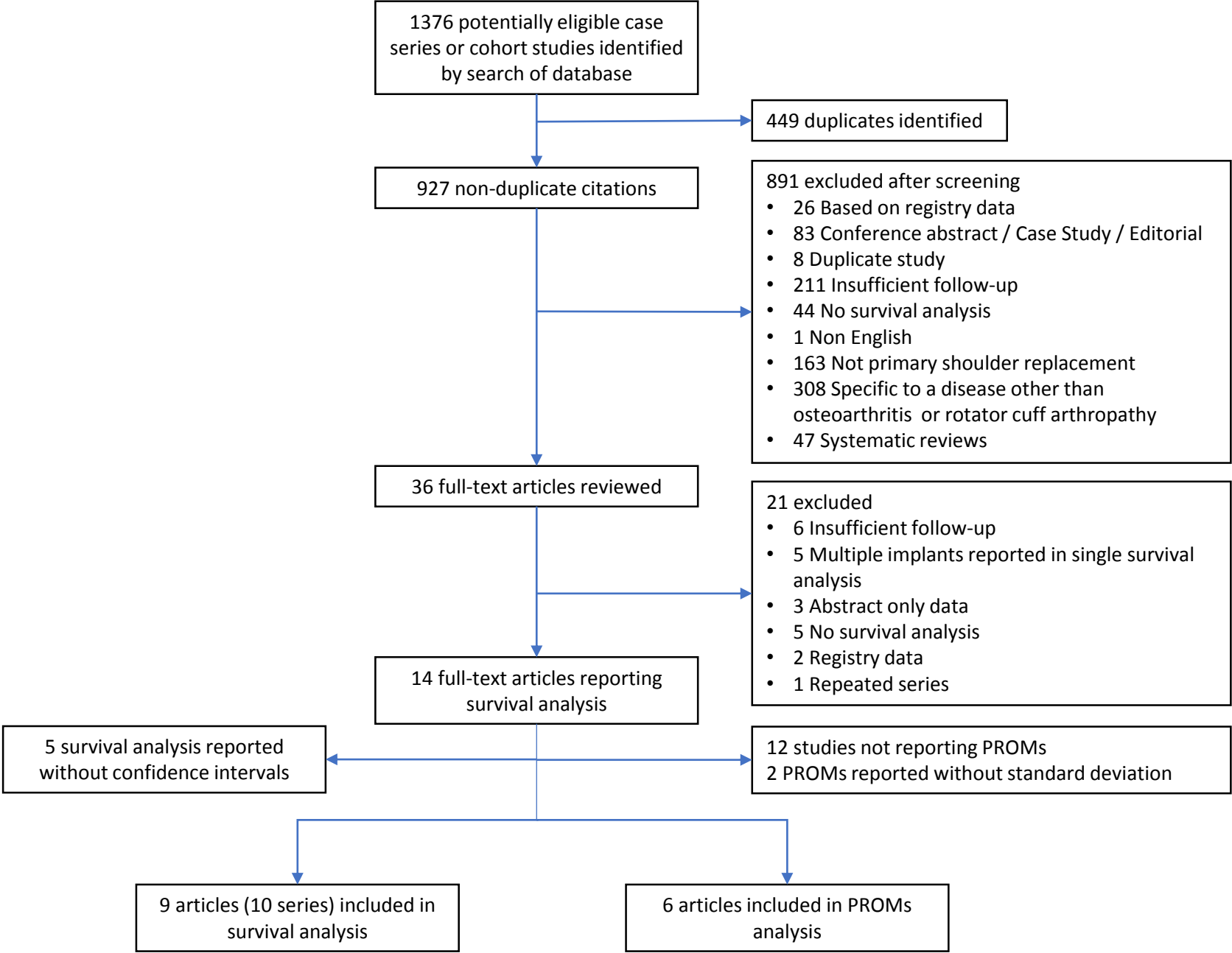
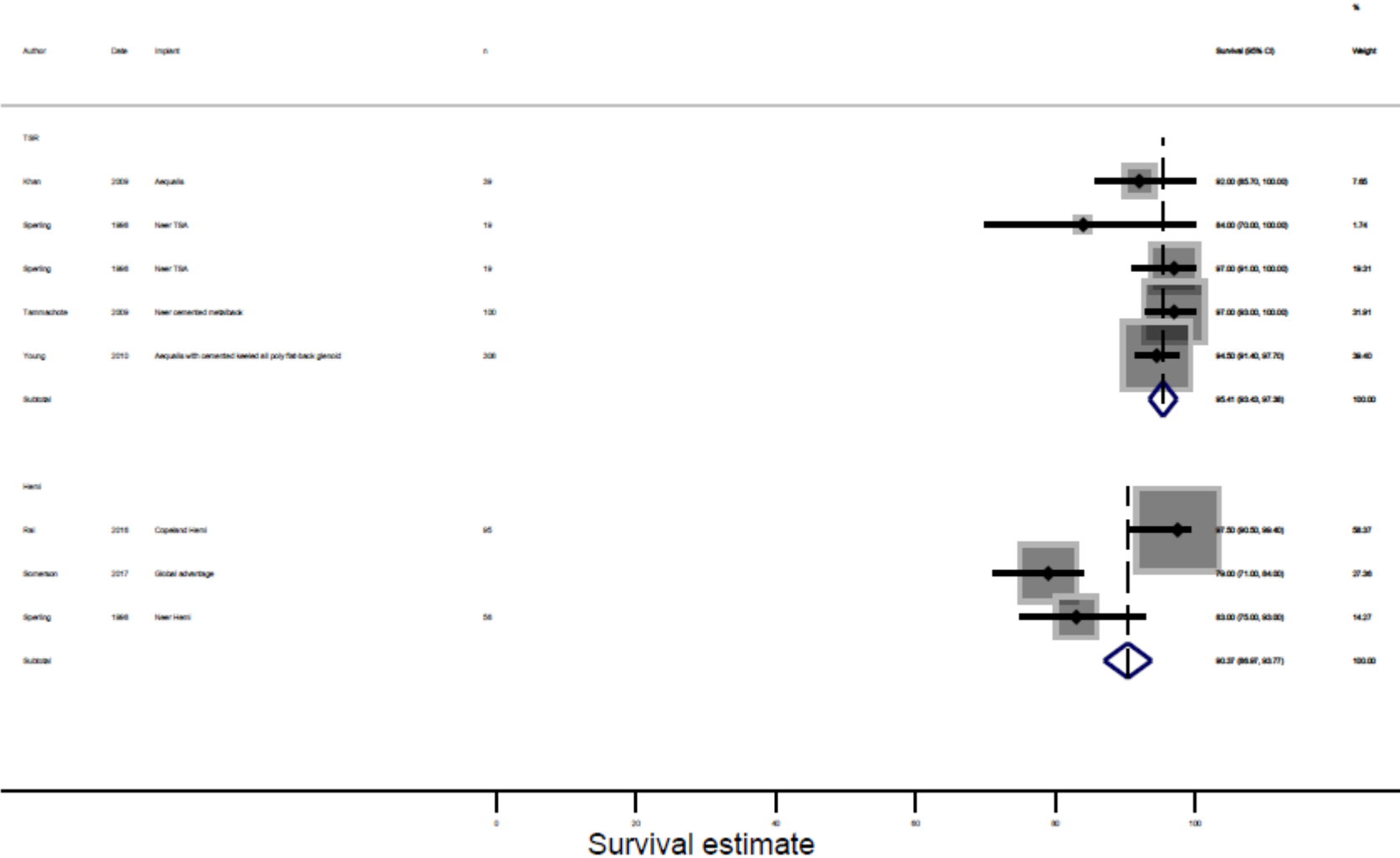
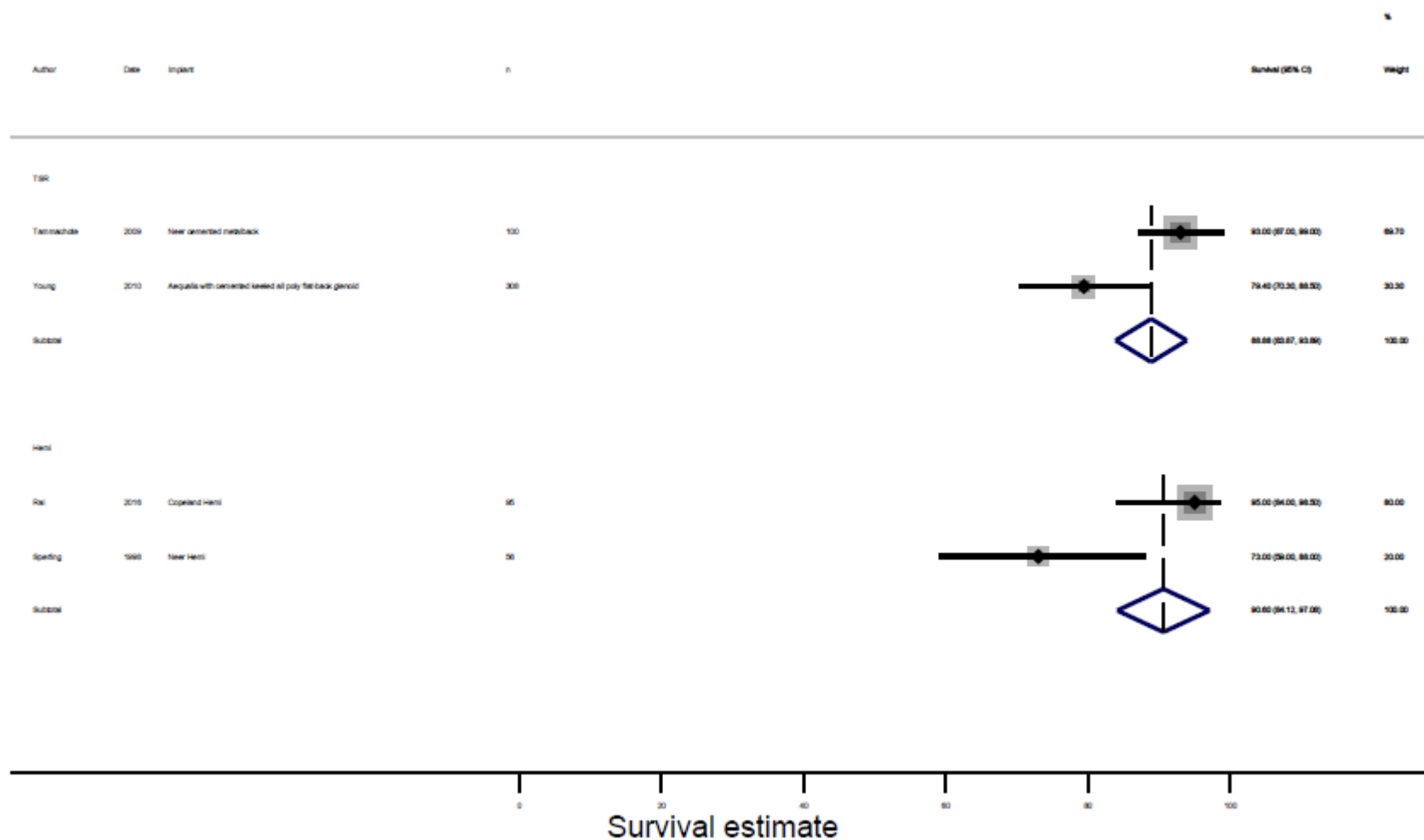


Figure 2

Figure 2: Forest plot estimates for survival of shoulder replacements from case series at 10 years (a), 15 years (b) and 20 (c) years



a



b

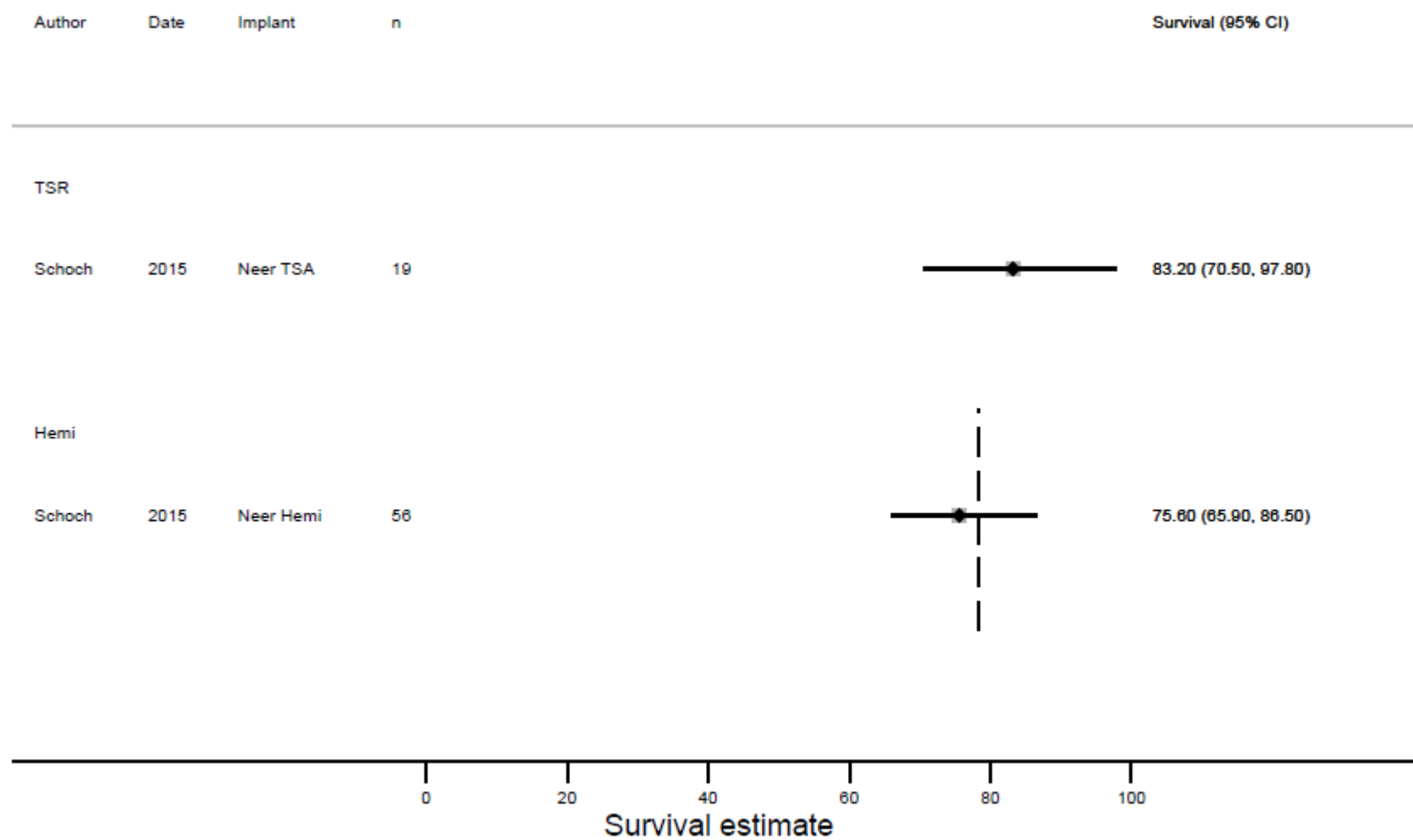
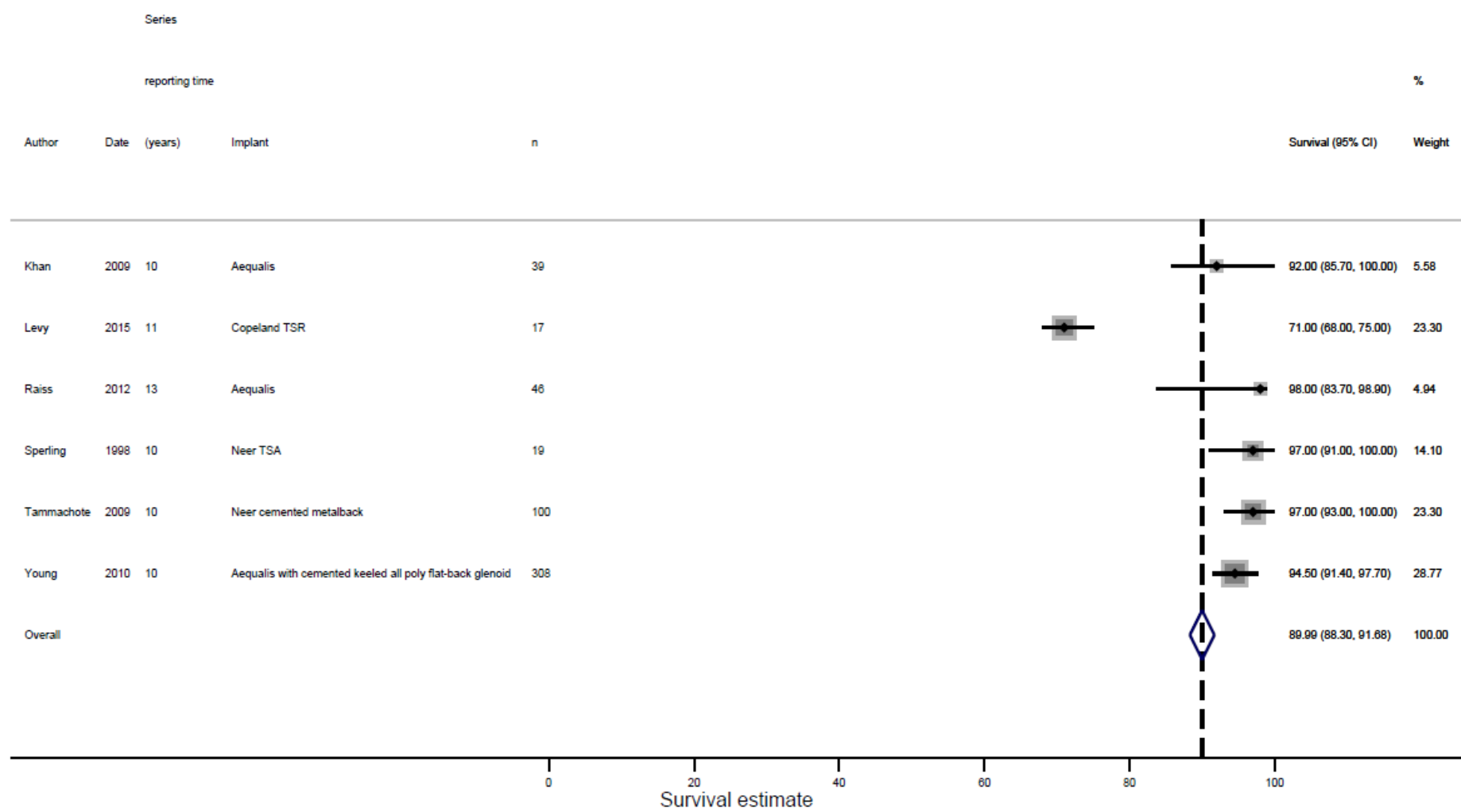
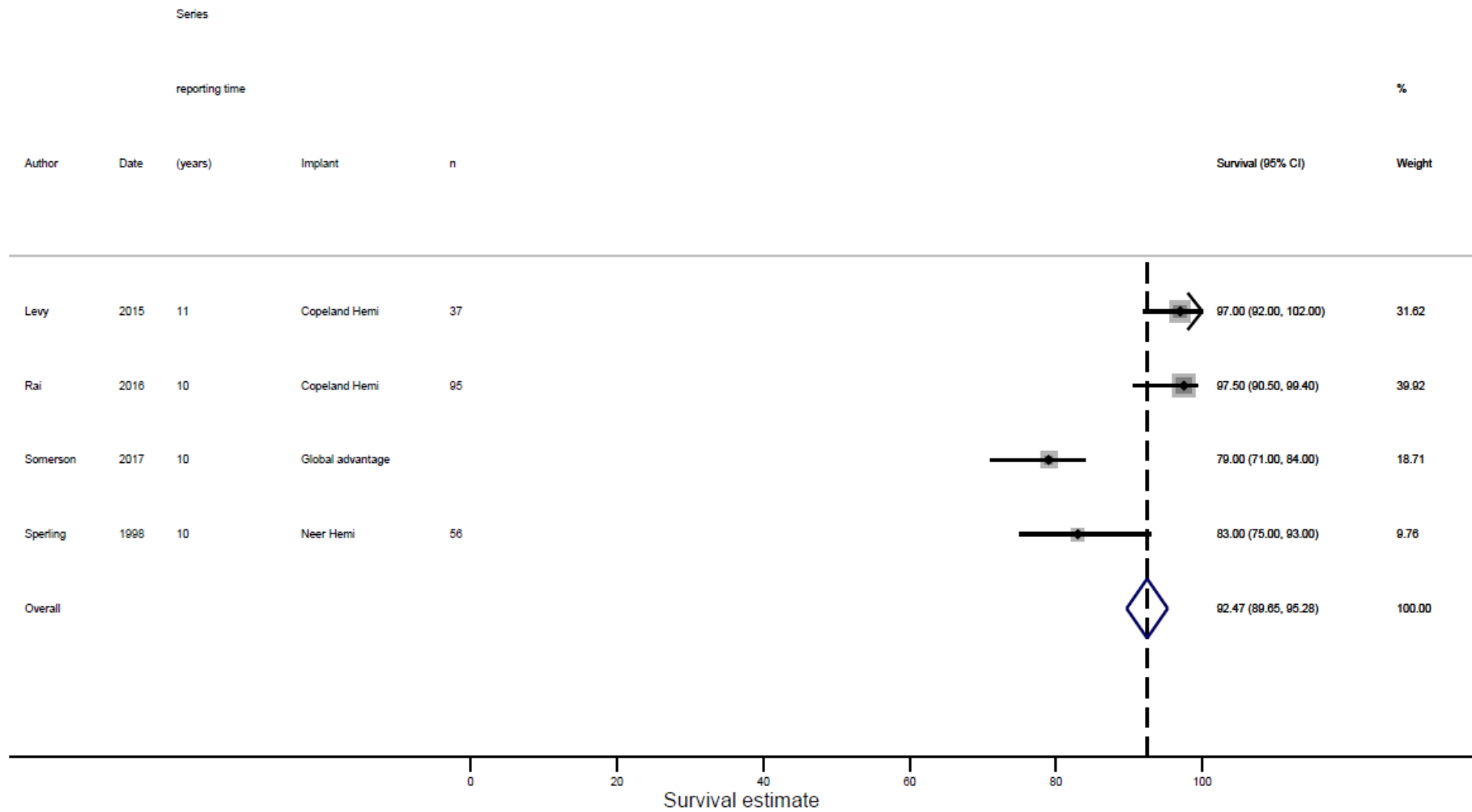


Figure 3

Figure 3: Forest plot of estimates of survival for reported survival of shoulder replacements from case series with rounding to 10 years. 3a Total Shoulder Arthroplasty, 3b Hemiarthroplasty



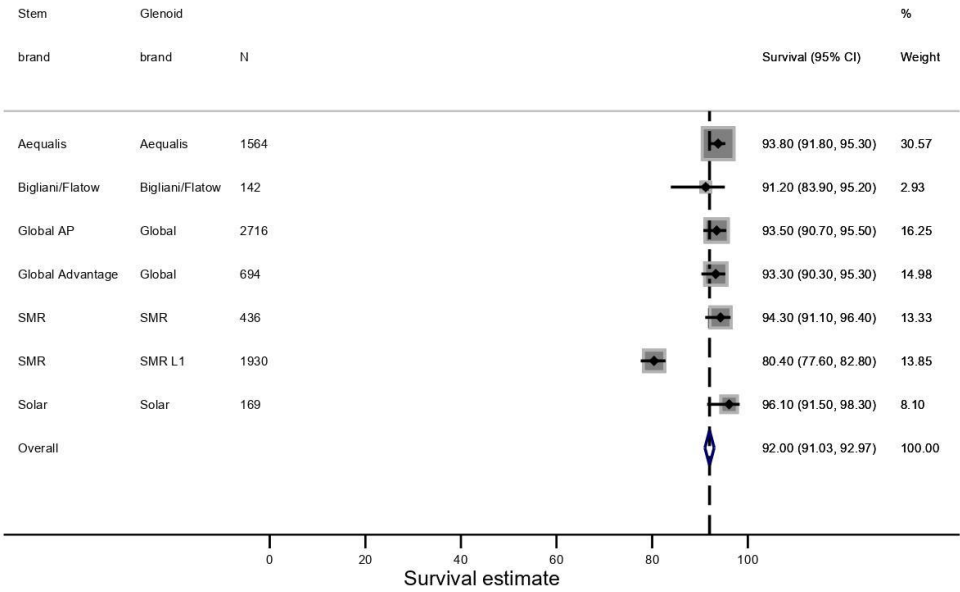
a

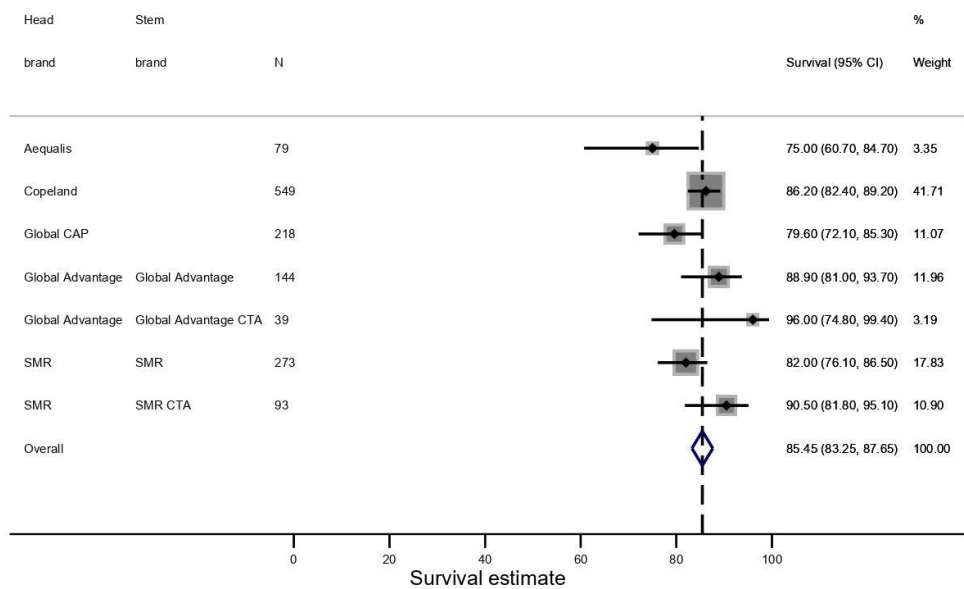


b

Figure 4

Figure 4: Forest plot of estimates for reported survival of shoulder replacements from registry reports at 10 years. 4a Total shoulder replacements, 4b hemiarthroplasty, 4c Reverse total shoulder replacement





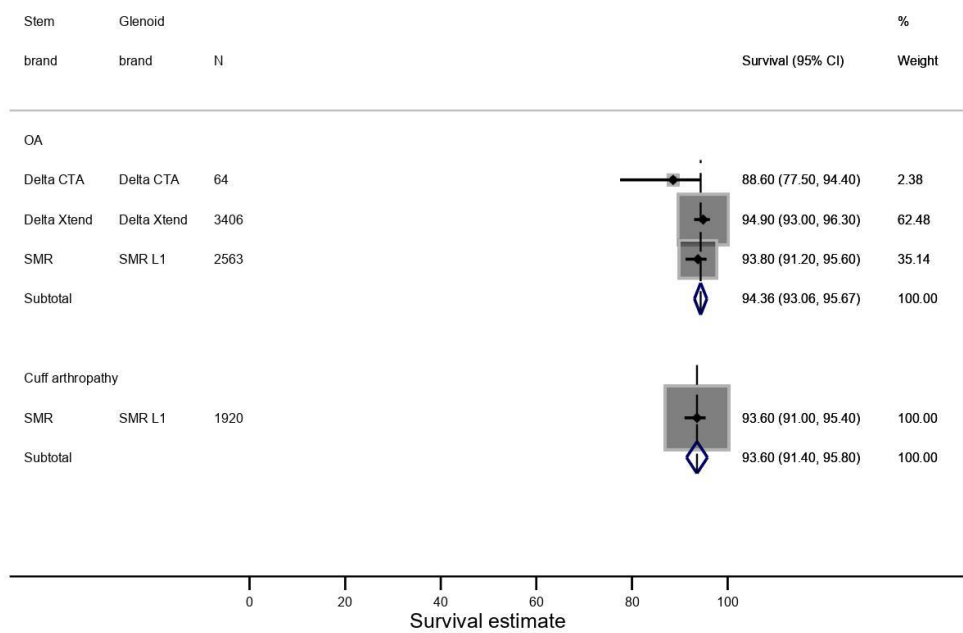
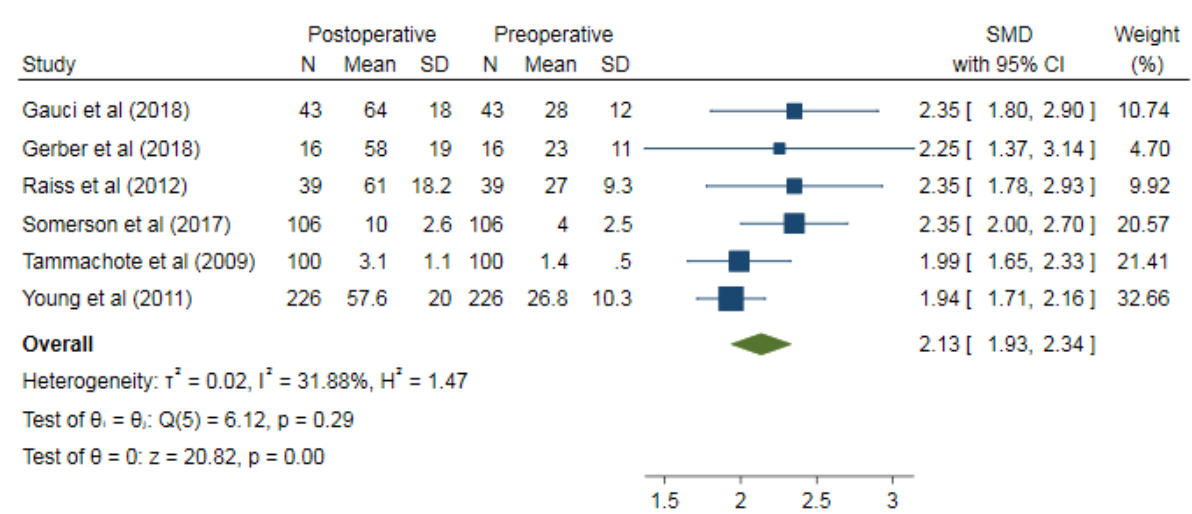
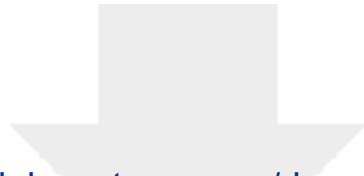


Figure 5

Figure 5: Forest plot of estimates of standardised mean difference (SMD) in Patient Reported Outcome Measures (PROMs) score following shoulder replacement at 10 years.





[Click here to access/download](#)

Necessary Additional Data

[How long does a shoulder - V2 Appendix 1 and 2.docx](#)



Response to reviewer's comments for the article titled:

How long does a shoulder replacement last? A systematic review and meta-analysis of case-series and national registry reports with more than 10 years of follow-up

We would like to thank all the reviewers for their very insightful and detailed reviews.

Please note, as some comments/questions had multiple stems, we have presented our responses in simple sequence rather than columnated format.

Comments are highlighted in red. The inserted / corrected text for the manuscript is shown within quotation marks. The associated page, paragraph and line numbers relate to the resubmitted manuscript with tracked changes.

All co-authors have contributed and read this document and the final manuscript

Editors' specific points:

1. Figures: Please supply editable files (eg, EPS files, PowerPoint files, depending on software used to produce them) for all figures such as diagrams and graphs. If figures are composed of photographs or other images, high-resolution files (300 dpi or greater) should be provided. If your figures are annotated, please supply two copies of each of these figures as separate jpgs (one copy annotated and one non-annotated). Our in-house illustrators will annotate according to journal style using the annotated figures as a guide. For multi-part figures, please supply the individual parts as well as a combined version to be used as a guide for our illustrators to recreate the files.

Thank you for highlighting this. We have checked all of our images and believe them to be compliant.

Figure 1 has been uploaded as a .pdf. Due to the software used we are limited to this or png/jpeg/svg. If this is not appropriate, we would be happy to redraw this figure within powerpoint.

2. Please complete and return a PRISMA checklist:

<https://eur03.safelinks.protection.outlook.com/?url=http%3A%2F%2Fwww.plosmedicine.org%2Farticle%2Finfo%253Adoi%252F10.1371%252Fjournal.pmed.1000097&data=02%7C01%7Cj.p.evans%2F40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023787104&data=2m6%2FETtVTvVJNt5Ls3zroGUZHPBnAFszMvTvOK4W9Fs%3D&reserved=0>

Thank you, we have now completed a PRISMA checklist and this is now attached.

3. Please check the PROSPERO registration number provided (CRD4201910221) as this number is not searchable on the PROSPERO website.

Thank you for highlighting this, we can confirm this was a typographical error and we have confirmed the correct registration number is CRD42019140221. This has been changed within the manuscript and we apologise for the error.

4. Lancet style is to have a 'Role of the funding source' at the end of the methods. The following points need to be addressed in the "Role of the funding source" statement:

- a) The role of the funders in the study design.
- b) The role of the funders in the collection, analysis, or interpretation of the data.
- c) The role of the funders in the writing of the report.
- d) Those who had access to the raw data (by author initials).

If the funding source had no role then this should be stated (ie, "The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report"). Please also add to this section (if true, or amend if not): "The corresponding author had full access to all of the data and the final responsibility to submit for publication."

Thank you for pointing out this omission. We have added the following text within the manuscript.

"Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had access to the raw data. The corresponding author had full access to all of the data and the final responsibility to submit for publication"

PAGE 8 /PARAGRAPH 3/ LINE 181 - 184

5. Please note that all references cited in the appendix (and all those included in the meta-analysis) must also be cited in the main paper.

We would be very happy to comply with this request, but due to the large quantity of papers included, this would result in either a reference list that exceeds the limit of 30, or has to omit references supporting the manuscripts content. We are therefore concerned that unsupported statements within the manuscript would have to be withdrawn, which would be detrimental to the reader. Could we respectfully request further guidance on this matter?

6. Please complete an authors' contribution and signature form; note that either hand-written or electronic signatures are accepted. Forms are available to be downloaded from <https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fels-jbs-prod->

cdn.literatumonline.com%2Fpb%2Fassets%2Fraw%2FLancet%2Fauthors%2Ftlrheum-author-signatures-1555082865647.pdf&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023797096&sdata=j1e3gazwABLyJPPsfQGtz%2BWoaXmzGL4VdnhewpjrUYw%3D&reserved=0 The contributions outlined on the form should match those in the text of the paper.

We have completed these forms and will forward them on as soon as possible

7. Please complete ICMJE conflict of interest forms (1 for each author); forms can be found at <https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.thelancet.com%2Ffor-authors%2Fforms%23icmje-coi&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023797096&sdata=yhcKejWwHuOANr77Es8bWF8IfiP%2FssA18sp%2BJpdRg0%3D&reserved=0>. The declarations on the forms should exactly match those stated in the declarations section of the paper.

We have complete this form and will forward it on as soon as possible.

8. Please supply the appendix as a single pdf file, with numbered pages. Please do not include a cover page with details of the paper, as we add a cover page to the appendix file before publication with this information

Thank you, we have now formatted the appendix as suggested.

9. As your research is funded by a body with an Open Access agreement in place with Elsevier (ie, by one of the Research Councils UK, Wellcome Trust, Cancer Research UK, Arthritis Research Council, British Heart Foundation, UK Department of Health, UK Chief Scientist Office, Austrian Science Fund, or Parkinson's UK), please consider now which licence you would opt for, should the paper be accepted for publication. There are two options - gold Open Access and green Open Access. Further details can be found at <https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.thelancet.com%2Fpb%2Fassets%2Fraw%2FLancet%2Fauthors%2Ftlrheum-info-for-authors.pdf&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023797096&sdata=4rXGyHfSDLxkXhKuyZm8ZnqsikV%2Bi427HS9lu09jLV4%3D&reserved=0>

We have discussed this amongst the authorship team, and under the current budgetary constraints, if accepted, would elect to pursue Green Open Access.

10. As your paper reports results of a systematic review and meta-analysis, please follow our formatting guidelines for these study types, available to download from <https://eur03.safelinks.protection.outlook.com/?url=http%3A%2F%2Fwww.thelancet.com%2Ffor-authors%2Fforms&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeeaf321334d8f04a53%7C0%7C1%7C637260993023797096&sd=ata=luLoRW1TGI%2FUcTXjuojtJhFWvZZ3zAwX8easPTBKCxE%3D&reserved=0> ('Systematic reviews and meta-analyses in The Lancet: formatting guidelines').

Thank you, we have reviewed the formatting guidelines and believe the manuscript is now in keeping with these.

11. Please check with your co-authors, and confirm that all names are spelt correctly, and affiliation details for each author are listed correctly (including department, institution, city, state [if applicable], and country). We cannot guarantee that we will be able to correct names and affiliations after publication of your article.

We have checked all of our co-author names and affiliation details as presented in the online upload and are happy that they are correct.

** **

Reviewers' comments:

Reviewer #1: Thank you for the opportunity to review this manuscript (How long does a shoulder replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 10 years of follow-up). The authors have done an excellent job evaluating the literature to identify the longevity of shoulder replacements. Notably, there is an exceptionally large volume of work evaluated, in what appears to be the first comprehensive pooled survival estimates for total shoulder arthroplasty. The findings from this study are important for patients and providers, especially as the demand for total shoulder arthroplasty continues to rise year over year. Although the manuscript is well thought-out, there are some revisions that need to be considered:

Thank you very much for these comments, we agree that this information is much needed by both patients and providers.

1) The Introduction should include specific data from other studies commenting on the rates of shoulder replacements performed globally or by country, as well as data highlighting their outcomes.

Thank you very much for this suggestion. We agree it is an important point however we were very limited by the word count in the original submission. We have followed your suggestion by further highlighting the work of Lübbeke et al (2017) in the opening paragraph (see below). We hope this gives the reader an impression of this rising incidence and geographical variation without using excessive text. Concisely summarising the outcomes following shoulder arthroplasty is, however, far more challenging (as we go on to demonstrate within this study) so we have elected not to include this information in the introduction.

“A review of seven national arthroplasty registers in 2017 suggested there has been a secular increase in the number of shoulder replacements performed for patients with both osteoarthritis and rotator cuff tear arthropathy. Overall, the annual incidence rate has increased 2.8 fold in the last decade, but significant variation exists between countries”

PAGE 4 / PARAGRAPH 1 / LINES 86 - 90

2) The end of the Introduction should clearly state the specific aims as they appear in the Results.

Thank you for highlighting this discrepancy to us. We have now amended the final paragraph of the introduction to keep it consistent with the results section.

“We aimed to provide the best quality pooled estimates of implant survival at a minimum 10 years’ follow-up. The decision to revise a poorly performing shoulder replacement is multifactorial that may be sensitive to both patient and surgeon preferences. Therefore, we also aimed to make a pooled estimate of the likely patient reported outcome at long-term follow-up, in essence to answer the question: Will my shoulder be better 10 years after surgery?”

PAGE 5 / PARAGRAPH 1 / LINES 109 – 112

3) What were the years of data included? "From commencement" needs to be made more clearly. If more recent studies were included, then that needs to be emphasized.

Thank you for this suggestion. We agree that making it easier for the reader to know the dates of origin of contributing articles is an improvement. As such we have now summarised this clearly in the results section.

“Six unique series, published between 1998 – 2015, reported survival of 529 total shoulder replacements (TSR) at 13 time points with follow-up ranging from 10 to 21 years (Appendix 2).”

PAGE 9 / PARAGRAPH 3 / LINES 200 - 201

“Four unique series, published between 1998 – 2017, reported survival of 364 shoulder humeral hemiarthroplasties (HAs) at 10 time points with follow-up ranging from 10 to 21 years (Appendix 2).”

4) The authors need to consider advances in shoulder arthroplasty (i.e. Grammont style prostheses, new techniques, etc.) that could influence recent vs. older outcomes.

We agree that advances in shoulder arthroplasty are likely to influence older results and this is an important point. New technology or change of technique could potentially lead to either superior or inferior results. Similarly, results could be influenced by changes in patient characteristics/demographics over time. We have now specifically highlighted this point in the text.

“We also recognise that emergent techniques and implants may demonstrate superior (or inferior) survivorship and function that is yet to be demonstrated with long-term follow-up.”

5) Can the authors comment on why these operations failed or why patients reported low outcomes?

This is an important question but unfortunately is beyond the scope of this study. Lack of detail and heterogeneity in the reporting of different series would make reliable synthesis of this evidence impossible. Furthermore, as we discuss in the limitations section, revision thresholds may be different dependent upon surgeon/centre. We would hope that correlations between PROMs and failure type/cause will be explored as the registries mature, but currently the present data is not clear on this.

6) To put it in context, consider including other/rates of other complications following shoulder replacements for comparison

We agree with this suggestion that other endpoints may be of interest as well as those reported in this study. Our chosen endpoint was guided by our patient group and we believe is the one of most interest at the present time. Furthermore, and importantly, the data reported in the included studies was variable both in terms of what was described and when/how it was reported. This heterogeneity would introduce bias to the study as well as potentially confuse the message of the article. Complications are certainly something we are keen to explore in future studies, but at the moment, is beyond the scope of this review, where we have used outcomes that have common definitions and reporting thresholds, are understandable and patient guided.

7) The Results section should be limited to describing the results of the earlier defined aims. It might be more appropriate to move lines 170 to 179 describing article selection to the Methods.

Thank you for this suggestion. Our structure is guided by the PRISMA checklist, which recommends the reporting of study selection numbers to be in the results section. We are certainly happy to move this at the discretion of the editor.

8) The inclusion of registry data needs to be made clearer. Was this registry data directly extracted by the authors from the registry as a separate analysis, or was this data already reported and the authors are extracting it from previously published work?

Thank you for highlighting this lack of clarity in the manuscript. Results from registries were directly extracted from registry annual reports and it was these data that were used for the registry analyses. The search terms did also identify several research articles reporting results of registry data. These articles were reviewed for references but results not included in data synthesis to avoid duplication of records. We have clarified this in the manuscript in two places.

“For inclusion, the case-series or published registry report had to report the survival of a specific brand of implant with a mean or median follow-up of greater than 10 years.”

PAGE 6 / PARAGRAPH 3 / LINE 130

“Reports were identified through the systematic search if published, or accessed through their websites.”

PAGE 7 / PARAGRAPH 3 / LINES 149 – 150

9) The authors need to better define "PROMs." What evaluations were utilized to make up PROMs (i.e. DASH, VAS, etc.)?

Thank you for this suggestion. We have now modified the methods to confirm that all PROM data was extracted and given examples (e.g. VAS, Constant, DASH) to guide the reader.

“Data extracted were: publication date, baseline population demographics, number of patients (n), surgical indication proportion (% OA and/or % RCTA), follow-up duration (>10 years), implant name and construct type (TSA, HA or RTSA), loss to follow-up, survival estimates (including CIs) and all available Patient Reported Outcome Measure (PROM) (e.g. Visual Analogue Scales (VAS), Constant score, Disabilities of the Arm, Shoulder and Hand (DASH)), data (outcome measure used, (outcome measure used baseline mean score (SD), follow-up duration in 5 year increments, follow-up mean score (SD)).”

PAGE 7 / PARAGRAPH 4 / LINES 155 - 161

10) It does not appear that the Figures are referenced in the text.

We have reviewed the manuscript again in full and ensured that all the figures are referred to in the text at appropriate places. We have changed the abbreviation “fig”, to “figure”. We have included an example of how we referenced the figures below.

“The search of published case-series produced 1,376 articles. Of these, 449 duplicates were removed, leaving 927 articles for screening (figure 1).”

ALTERED THROUGHOUT THE RESULTS SECTION

Reviewer #2 (Statistical review): TLRHEU-D-20-00161 How long does a shoulder replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 10 years follow up

The authors present a well written article on an important issue with limited clinical or patient data analysed to date, and report what seem clear and useful findings. The methods for the identification and extraction of the relevant data in the systematic review look appropriate, likewise the statistical methods for the meta-analysis seem expertly applied. There is an adequate level of supporting data given via the tables and usual forest plots. There are several presentational and statistical issues to address, as follows:

Thank you very much for your kind comments.

1. The authors state that (line 127) 'For inclusion, the case series had to report the survival of a specific brand of implant with a mean or median follow up of greater than 10 years'. This needs to be justified in more detail:

Thank you for requesting further clarification on this. We required a specific brand of implant as studies on hip and knee replacements have shown different implants to have different survival results. It is logical to assume the same is true of shoulder arthroplasty, and therefore limited included articles to those reporting results for each single brand that was in the study. The following text in the manuscript discusses this.

“Although this has not specifically been assessed in shoulder replacements, the technique of treating each brand as its own series was utilised as variation in survival by brand exists in hip and knee replacements, therefore the assumption would seem sensible for shoulder replacements as well. Weighting of implants in the meta-analysis would therefore provide the most robust survival estimates.”

A cut off of 10 years was chosen to give the required combination of enough evidence to allow robust analysis without having excessive numbers of articles to include (which would likely be the case if we included all articles reporting survival). The threshold in this study was 10 years whereas our previous studies of hip and knee replacement used a cut-off of 15 years which reflects the lower number of articles reporting long-term survival of shoulder replacements. We have further clarified this in the manuscript.

“A cut-off of minimum mean or median follow-up of 10 years was chosen as the subject of interest of this study was “long-term” survival, where there is a current paucity of information. We accept this definition may vary subjectively but 10 years allowed inclusion of sufficient studies to make analyses robust and represents a time period that is relatable to patients and clinicians.”

a. They also state line 149 that 'Data were not extracted from figures (e.g. Kaplan Meier plots) to avoid potential transcription inaccuracy'. Yes, that is a potential problem, but there are simple digital apps available that minimise the inaccuracy.

Thank you for this helpful suggestion. This is something we considered and have used in previous research. Unfortunately, this requires an “at risk” table which in the articles identified was almost never included. In addition to this, some of the articles were quite old with very poor-quality images that may give inaccurate figures even if an “at risk” table was included. This approach also kept this study consistent with the methods for our previous studies of hip and knee replacements published in the Lancet.

b. By requiring the mean or median to be >10 years, they do focus on these series with longest term follow up - but they acknowledge the other problem that the longer the follow up, generally the higher the level of loss to follow up i.e. missing data. Series with <10 years follow up mean or median can still be informative about 10-year survival rates, since the survival rates they are using will all have been calculated under censoring (albeit assuming that censoring was non-informative).

Thank you for highlighting this, this is something we considered at length when the study was in the design stage. This is particularly relevant to registry studies which will have a high volume of patients with longer follow-up but potentially a low mean follow-up, due to the increasing numbers of primary arthroplasties being performed each year.

We agree that in properly conducted analyses, censoring will mean that 10-year survival estimates may be reliable with a mean/median follow-up of under 10 years. Our experience of survival reports

within the orthopaedic field however tells us that authors often report survival at the longest time possible, often when only one patient remains at risk. An article by Lettin et al¹, suggests that in hip arthroplasty a minimum of 40 joints at risk is required for accurate survival estimates and it is reasonable to suggest a similar cut-off would be the case in shoulder arthroplasty. The poor reporting of the number of joint replacements at risk at each reporting time means that we would not be able to ensure the quality of survival estimates. In view of this, to maximise the quality of evidence, we took the decision to only include case-series with a mean/median follow-up of 10 years or more. Alternatively, we could synthesise pseudo-individualised patient data from Kaplan-Meier graphs to account for this censoring, however as stated above there were few papers that included “at risk” tables to allow us to do this. We do suggest in the manuscript that inclusion of “at risk” tables with all Kaplan-Meier graphs should form a basic reporting standard.

“Although component-failure data are of interest, we would regard this as best reported as a secondary endpoint, with the all-cause 1-Kaplan Meier estimate as the most appropriate method of reporting survivorship, which should always include the number of shoulder replacements at risk at the time of reporting.”

PAGE 13 / PARAGRAPH 3 / LINES 288 - 291

1. Lettin AW, Ware HS, Morris RW. Survivorship analysis and confidence intervals. An assessment with reference to the Stanmore total knee replacement. The Journal of bone and joint surgery British volume 1991; 73(5): 729-31.

c. So, although the approach the authors have taken is understandable, in that it might (or should) minimise some biases, unintentionally it may have allowed other biases to be larger. Did they look into dropping this inclusion criteria to e.g. 9 or 8 or 7 years, say?

Thank you for this observation. We agree with this theory, however we aimed to find the appropriate balance of potential sources of bias. This combined with the other reasons described above, we chose to set our criteria for a mean/median follow-up of 10 years.

d. And having raised the issue of loss to follow up, what was this like across the included studies - there is only mention of <80% being an exclusion criterion?

Thank you for this comment. Loss to follow-up of <80% formed part of the quality assessment, and was not an exclusion criterion. We agree that the range of number of patients lost to follow up would be an interesting inclusion in the manuscript so have now included the mean number and range of patients lost to follow-up.

“Quality assessment revealed that six (60%) of the 10 series were consecutive, two (20%) were multicenter, nine (90%) had >80% follow-up (with mean loss to follow up of 8.4%, ranging from 0%

to 23.7%), and none undertook multivariable analysis. These proportions are in keeping with the fact that the quality of published case-series is low.”

PAGE 9 / PARAGRAPH 2 / LINE 195 - 198

2. The authors state that 'Given time, all shoulder replacements will fail' - could the authors indicate what this natural failure distribution is like?

This statement relates to the fact that all arthroplasties (hip, knee, shoulder etc) will fail eventually if left in situ long enough. The natural failure distribution of shoulder replacements is unclear to the best of the authors knowledge and is one of the reasons for this study. In response to your comment we have changed the manuscript to avoid potential mis-understanding.

“Some shoulder replacements will fail and further operations may be required”

PAGE 3 / PARAGRAPH 1 / LINE 58

3. The article reports on TSR, HA, and RTSR - it would be useful, given the decades long time horizons being discussed, if the authors could indicate the temporal trends in the use of these procedures, and the characteristics of the patients receiving these (and the sites / surgeons offering them).

Thank you very much for this suggestion, which was echoed by your co-reviewers. We have amended the introductory paragraph to highlight the temporal trends in use and geographical variation (see below) which we hope gives the reader an impression of the rising use and classes of associated conditions. However, concisely summarising the patient characteristics would prove very challenging, as there is large heterogeneity both at the patient and surgical application level. We feel that it is the ability to have an individualised patient-centred discussion about surgical options that is paramount, and it's this discussion that we hope to facilitate with our findings. As you will have read, we do go on to discuss the rising incidence in RTSR use, which contrasts a paucity in evidence, but feel that this is better placed within the discussion,

“A review of seven national arthroplasty registers in 2017 suggested there has been a secular increase in the number of shoulder replacements performed for patients with both osteoarthritis and rotator cuff tear arthropathy. Overall, the annual incidence rate has increased 2.8 fold in the last decade, but significant variation exists between countries”

PAGE 4 / PARAGRAPH 1 / LINES 86 – 90

a. These temporal trends might be influenced by new brands or technologies, and perhaps increased confidence in using them by a wider pool of surgeons, with less challenging learning curves - and so on.

Thank you for this insightful observation. This is indeed possible; however, this study is not seeking to analyse temporal trends in survival. This analysis would need much greater numbers and

granularity than was identified in our review. Given the paucity of evidence regarding long term survival of shoulder replacements we tried to keep the message simple but valuable. We do discuss the increasing use of RTSR which is currently the preeminent technological shift and as we discuss, without much supporting long-term survival evidence. It is perceived that the learning curve is steeper for this technology, but our data suggest that the survival at 10 years is better than TSR and HA. This may however be influenced by the fact that there are few (if any) options for revising a “failed” RTSR.

b. It would be useful to understand the dynamics of the context that these data were accrued in? The authors just state (line 85) that 'There is a secular increase in the number of shoulder replacements performed for patients with osteoarthritis and rotator cuff arthropathy'?

Thank you for highlighting this. In hindsight we had not supported this statement or put it into context. Within the same paragraph we highlight above, we have adapted the manuscript now to offer some context to this statement.

“A review of seven national arthroplasty registers in 2017 suggested there has been a secular increase in the number of shoulder replacements performed for patients with both osteoarthritis and rotator cuff tear arthropathy. Overall the annual incidence rate has increased 2.8 fold in the last decade, but significant variation exists between countries”

PAGE 4 / PARAGRAPH 1 / LINES 86 - 90

4. It was confusing that although the authors seemed to identify only one registry (the Australian) they mention multiple 'series' for each of the 3 operations within this registry - and then go on to report the 'pooled registry 10-years survival' rates. So, what are these series (e.g. one series for each specific brand, say?) and is then the 'pooled' estimates then pooling across these brands?

Yes, you are correct in your interpretation. We believe that the implant itself is fundamental to the survival outcome of surgery and therefore we chose the implant construct used as our exposure of interest. In view of this, each individual series constitutes and should be considered as a different patient cohort. Although all the individual series estimates were from a single national registry, we were able to retrieve 11 different survival estimates for HA at 10 years, 7 for TSR and four for RTSR.

We have used these estimates to synthesise a single pooled estimate, weighting the estimates according to standard error. This type of analysis, that derives an overall estimate according to how frequently each implant has been used, is not provided in the Australian registry and is unique to our study.

As this point is not clear to the reviewer, we accept that it may not be entirely obvious to the readership, so have added an explanation into the discussion.

“The implant has been shown to be fundamental to the survival outcome of hip and knee replacement and is likely to be just as important in shoulder replacement and each individual series should be considered as a different patient cohort. We have used the individual estimates for each implant to synthesise a single pooled estimate, weighting the estimates according to standard error. This type of analysis, deriving an overall estimate according to how frequently each implant has been used, is unique to our study.”

PAGE 12 / PARAGRAPH 3 / LINES 252 - 257

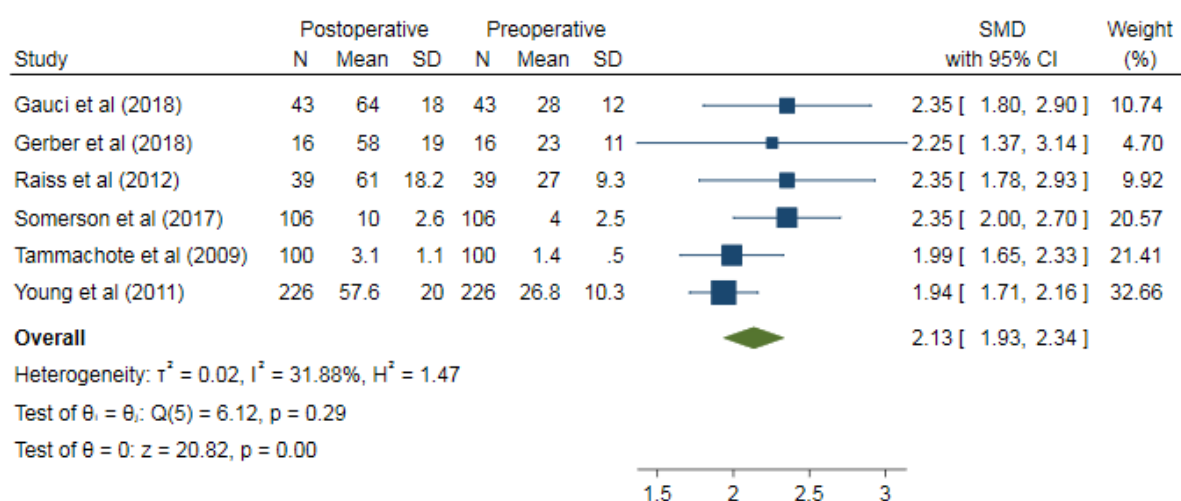
5. The authors report that 'Pooled 10-year PROMs revealed a substantial improvement over baseline scores' - so does this mean that there was always a baseline score, and was this always measured before the operation?

Yes you are correct, all the included articles reporting PROMs scores had a baseline score that was taken pre-operatively. These are represented in the PROMs forest plot which we have added below.

“Data extracted were: publication date, baseline population demographics, number of patients (n), surgical indication proportion (% OA and/or % RCTA), follow-up duration (>10 years), implant name and construct type (TSA, HA or RTSA), loss to follow-up, survival estimates (including CIs) and all available Patient Reported Outcome Measure (PROM) (e.g. Visual Analogue Scales (VAS), Constant score, Disabilities of the Arm, Shoulder and Hand (DASH)), data (outcome measure used, (outcome measure used baseline mean score (SD), follow-up duration in 5 year increments, follow-up mean score (SD)).”

PAGE 7 / PARAGRAPH 4 / LINES 155 - 161

Figure 5: Forest plot of estimates of standardised mean difference (SMD) in Patient Reported Outcome Measures (PROMs) score following shoulder replacement at 10 years.



6. Line 129 'Although this has not specifically been assessed in shoulder replacements, this methodology was utilised as weighting of implants in the meta-analysis provides the most robust survival estimates' - it is not at all clear what the authors mean by this, or what this 'methodology' (analysing by brand?) actually refers to?

We apologise for the lack of clarity. We refer to the fact that we treat each individual brand series independently and weight the impact accordingly. The methodology we refer to was that used in the previous papers regarding hip and knee replacements published in the Lancet. We have clarified this in the manuscript.

“Although this has not specifically been assessed in shoulder replacements, the technique of treating each brand as its own series was utilised, as the assumption that variation in survival by brand that exists in hip and knee replacements would seem sensible for shoulder replacements as well. Weighting of implants in the meta-analysis would therefore provide the most robust survival estimates.”

PAGE 6 / PARAGRAPH 3 / LINES 132 – 136

7. Line 136 'systematic reviews were assessed for their citations but did not include their pooled data to avoid duplication' - that looks good - but what about any case series of Australian data and the Australian registry?

Thank you for your concern, we had the same thought, however on checking the Australian data is not reported in any published series, they only produce this data as part of their registry report, this has been confirmed with personal correspondence.

8. Line 157 and Line 293 - I wasn't sure what the authors meant by 'Survival estimates, assuming that survivorship approximated revision risk ...'; then Line 293 expands on this 'We also assumed that survival estimates are equivalent to risks ... and although the assumption that no censoring (patients dying with a shoulder in situ) is violated ...' - but that seems a major violation, given that e.g. the age of the TSR in the Australian registry was 72 so by 10 years follow up on average they will be 82? Was it possible to look at the hip data to see what the influence of this assumption might be (where there is individual level data)?

Yes, we agree, it is a potential flaw of this meta-analysis technique which is why we chose to highlight it clearly. We do however feel that although not perfect it is a reasonable method given the lack of alternative options to meta-analyse survival data. We assumed that within each class of implant that censoring will be similar across all cohorts which will minimise any adverse impact.

9. Line 162 there is a description of the study quality assessment, which is fine - but it wasn't clear if anything more was done with this - i.e. in terms of assessing the influence of study quality on the outcomes?

Thank you for asking us to clarify this. The results of the quality assessment are reported as general observations (see below). This bespoke tool derived by Wylde et al for use in musculoskeletal evidence synthesis is used to give a general idea of the strength of evidence rather than as a formal analytical tool. Sensitivity analysis dependent upon study quality was not part of our analysis plan, but as we have previously reported in the hip and knee long term outcome papers, the general quality of case series appears to be poor.

“Quality assessment revealed that six (60%) of the 10 series were consecutive, two (20%) were multicentre, nine (90%) had >80% follow-up (with mean loss to follow up of 8.4%, ranging from 0% to 23.7%), and none undertook multivariable analysis. These proportions are in keeping with the fact that the quality of published case-series is low.”

PAGE 9 / PARAGRAPH 2 / LINES 195 - 198

10. Lines 184 and 187 (and again, lines 191 and 193) - it wasn't completely clear what the authors had done with the 'exactly 10 years' (e.g. 95.6%) and the 'rounding down of reported survival between >10 and <15 years' (e.g. 90%) - could this be explained more carefully?

Thank you for highlighting this. We agree the text did not make it completely obvious and as such we have now modified this sentence and hopefully this now reads more clearly.

“When studies reported survival estimates at between 10 and 15 years, these results were rounded down to 10 years as a sensitivity analysis. This resulted in a pooled survival of six series (529 TSRs) of 90.0% (95% CI 88.3, 91.7) (figure 3).”

PAGE 9 / PARAGRAPH 3 / LINES 205 - 208

11. It was confusing to read about the 'New Zealand registry' (line 213) after the authors had stated only one registry (the Australian) had data that qualified. Presumably the NZ one qualifies for the PROMs but not the clinical outcomes (which seems a bit strange)?

Thank you for asking us to clarify this. We have tried to do this in the manuscript. The New Zealand registry was not included in the pooling of data because it did not report either survival or PROMs with sufficient detail for us to meta-analyse the data. Results were only reported by class of implant (HA, TSR, RTSR) rather than by brand of implant which was a requirement for inclusion in our study. With regard to PROMS, they also do not report baseline pre-operative scores and therefore an assessment of effect size cannot be made. It is however the only registry to report any results of 10 year PROMs and is therefore valuable as our only comparator for discussion.

“The New Zealand registry provided the only comparator of construct-level, but not implant level, PROMs data. At 10 years this was limited to 674 replacements. Their high OSS at 10 years (80% of total score) does suggest a sustained benefit of shoulder replacements. As the New Zealand registry does not provide baseline pre-operative scores, comparison of SMD could not be undertaken.”

PAGE 14 / PARAGRAPH 2 / LINES 302 - 305

12. Lines 251-254 - this is clearly understood, but wouldn't it have been informative to try to include these excluded data by way of a sensitivity type analysis, using appropriate techniques to recover/impute the missing data?

Thank you. We were very concerned about the missing data but we have adhered to our analysis plan. Our concern with using estimated CI data we have discussed earlier in point 1a. With regards to imputation of missing data, it is important to remember that this is not an individual patient data meta-analysis. The individual cohorts do not have or do not report missing data, but the authors of the included registry have combined cohorts of different implants, which as we discuss above are likely to have failed at different rates. There is no way that we are aware of to reliably untangle this data into implant cohorts. The principle of this review would therefore be violated if we analysed this data and we feel that the ability to impart a simple message would become challenging.

13. Figures 2-4abc don't include the usual indicators of / tests for heterogeneity - whereas Figure 5 does? Naively it would seem important to have these statistics in all the plots and sub-plots?

We thank the reviewer for this comment. In typical meta-analyses of treatment A vs treatment B, funnel plots as proposed by Sterne and Egger are useful methods for detecting publication bias as they are centred around a treatment difference. In this study, there is no treatment difference, and whilst it is possible to present a funnel plot which is centred around the fixed effect estimate from the metanalysis, it is less meaningful than would be observed in clinical trials data. Therefore, we believe that simply presenting the data and viewing the distribution is probably the best method of detecting publication bias.

Reviewer #3: The authors address an important topic: the "lasting" of a shoulder replacement.

Although shoulder implants are not used as regular as hip and knee implants, an increase of the use of these implants is apparent in many countries, and perhaps an overuse of the reversed type of shoulder implants. Related to the low frequency, data of meta-analysis and registry studies is necessary for clinicians who have patients with shoulder complaints. The latter is not only the orthopaedic surgeon, but also GP, who are the gate-keeper in many countries before referrals.

Thank you, we agree that it is useful information with all involved in the care pathway.

Some comments on the article:

Title

change ... "national registry reports" into "a national registry report with more than.... etc". This since only the Australian registry could be used

Thank you for highlighting this. Respectfully, we would recommend keeping the title unchanged as we believe the title to be appropriate and importantly the title as we have it keeps the article in line with the others in the series published by the Lancet. Given only one registry report presents 10-year survival results broken down by implant, this study does encompass all reports.

line 146 indication proportions (% OA and / or% RCTA) line 147 loss to follow-up

Proportion are outlined in:

PAGE 9 / PARAGRAPH 1 / LINES 192 – 194.

Loss to follow-up is now reported in:

PAGE 9 / PARAGRAPH 2 / LINE 196.

Some bias might occur by excluding studies with no CI (flowchart, 5 studies were excluded, thus in contradiction to line 297 of the discussion), since data are not missing at random. Research groups not reporting CI may do so routinely. These researchers may "on average" find different mean values. Even more, techniques exist to calculate CI. See these two references:

<https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fpubmed%2Farticles%2FPMC5796634%2Fpdf%2Femss-75934.pdf&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023797096&sdata=TmA%2F%2B7Og3mt2z7Hk8UwvbDSwwqEoqreFpH6KgnLblyc%3D&reserved=0>

and

<https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fsystematicreviewsjournal.biomedcentral.com%2Ftrack%2Fpdf%2F10.1186%2F2046-4053-3-151&data=02%7C01%7Cj.p.evans2%40exeter.ac.uk%7C0fc9509d98ef474f981208d8017f2c8f%7C912a5d77fb984eeef321334d8f04a53%7C0%7C1%7C637260993023797096&sdata=VqM44R1d4RaZNPEQO3biHAa%2BJvmMtLMbvjWfzGGJ6TM%3D&reserved=0>

If this would not work, the se/sds of comparable studies could be used as an estimate. The could also perform a sensitivity analyses, since they know from studies they use the se. Anyhow, just excluding studies will create more "uncertainty" about the found estimates of outcome.

Q to authors: could either method be done with your data, in order to have a more robust estimate of your outcome data.

Thank you very much for your suggestion of techniques. We were aware of these methods however did not feel them appropriate for use in this particular setting. Pseudo individual patient data meta-analysis is useful technique when KM curves are published with sufficient resolution and an "at risk" table. The failure to publish at risk tables limits these approaches. As "at risk" tables are rarely reported in this context we approximated survival estimates as rates and used standard methods.

Results

Table 1 Include loss to follow-up TSR, HA, RTSR.

Thank you very much for this useful suggestion. We have reformatted the table as seen below.

*Table 1: Study-level and participant-level characteristics of contributing data sources *estimates from whole group data. ‡weighted age and %female dependent on number of arthroplasties in the series. TSR – Total Shoulder Replacement, HA – Hemiarthroplasty, RTSR – Reverse Total Shoulder Replacement*

	Individual case series articles			Australian Orthopaedic Association National Joint Replacement Registry annual report 2019		
	TSR	HA	RTSR	TSR	HA	RTSR
Study level characteristics						
Location	UK (2), Germany (1), USA (2), Pan European (1)	UK (2) USA (3)	0	Australia	Australia	Australia
Number of unique implant series included	6	4	0	8	8	5
Year of publication	1998-2015	1998 - 2017	NA	2019	2019	2019
Participant level characteristics						
Mean age (years)	66.9 [‡]	54.5 [‡]	NA	72.2*	67.7*	74.3*
% of female patients	57.8 [‡]	34.4 [‡]	NA	61.6*	63.4*	64.3*
Total Arthroplasties (n) at start	529	420		7941	3495	8049
Loss to follow-up	62 (11.7%)	30 (7.1%)				

The HA series seems to have some bias? by an over representation of men? please comment in discussion.

Thank you for highlighting this important point. We have added a discussion of this point into the manuscript now.

“It is notable that the demographic characteristics from the case-series and registry data are similar for the TSR group, and concordantly their survival rates are also comparable. For the HA group, the case-series data contain a more male dominated and younger population. All but one of the case-series report an average age of <60yrs, therefore the survival findings from case-series may lack generalisability.”

PAGE 13 / PARAGRAPH 1 / LINES 270 - 274

What is coverage Australian Registry (100%?) completeness of primary and revision shoulder surgery of TSR. HA, RTSR

This is an important point, thank you. The Australian Orthopaedic Association National Joint Replacement Register (AOANJRR) report on their website that initial pass of validation against Health Department data revealed over 94% accuracy. Source: <https://aoanjrr.sahmri.com/data>

Figures 2, 3 mention "N" in a column and % OA ; % Fracture ; % RCTA (the latter will be "null", I would guess for the no RTSR)

Thank you for this suggestion. As stated in the methods, we assessed %OA / %RCTA and this is now reported in the manuscript (see below). However, due to the poor reporting of this data in the included articles we feel that it's inclusion to the forest plots will not be helpful to the reader. We have now included "n" in figures 2 and 3 as you suggest, to keep the graphs consistent (example below).

“The proportion of OA as the primary surgical indication was 59% for TSR and 48% for HA. The reporting of indication was variable and was interpretable in only seven articles.”

PAGE 9 / PARAGRAPH 1 / LINES 192 - 194

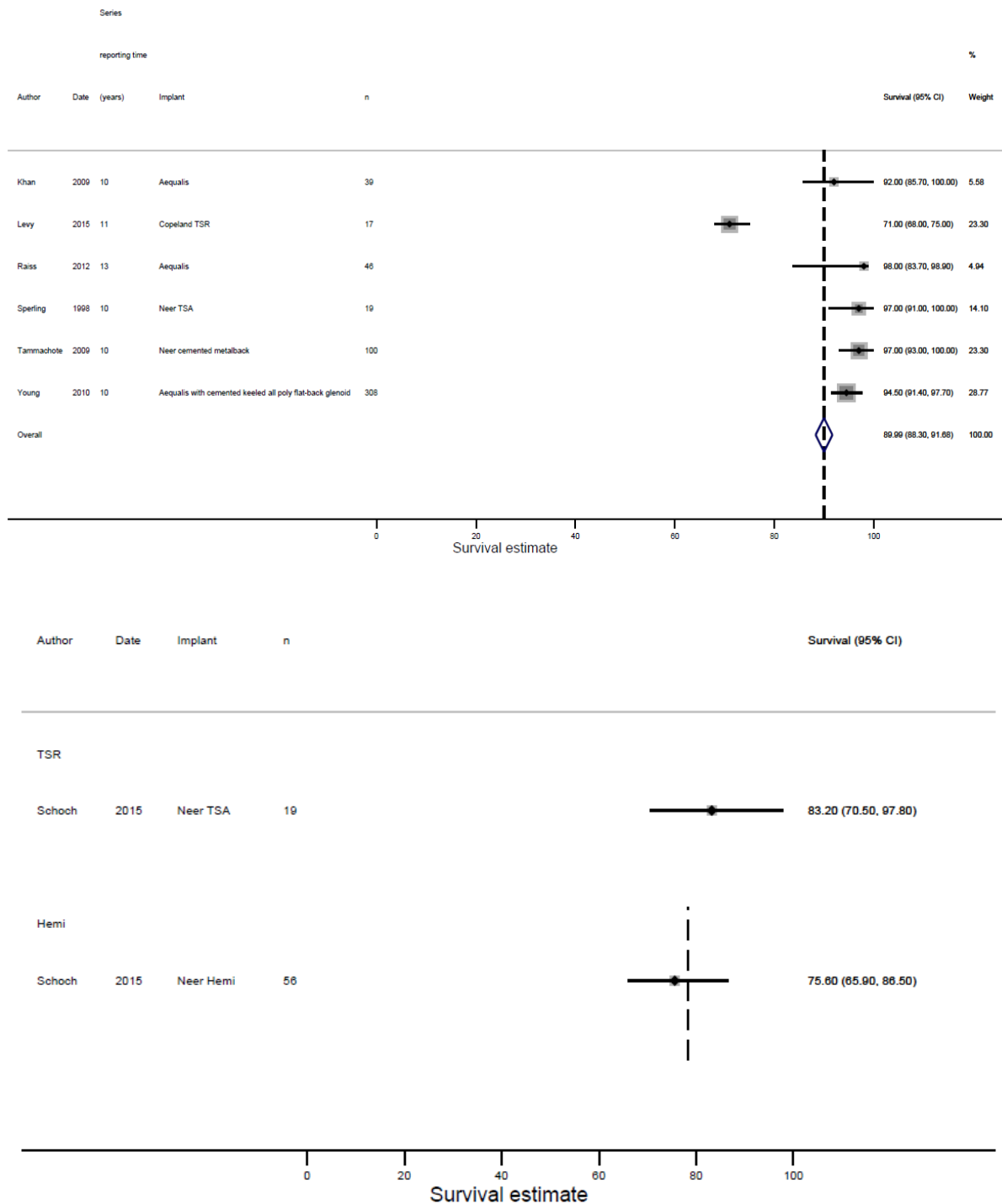


Figure 3, the study of Levy is an outlier, address this better in the Discussion section. Since in all studies the glenoid component may be the "weak" link, in the Copeland shoulder it was the humeral component?

This has been addressed in the limitations section. As Levy et al stated in their paper, they were using a metal-backed glenoid that had a high failure rate.

Figures 2,3,4, please mention like in figure 5 the statistics for heterogeneity, test etc

Thank you for this suggestion. As we discuss above in typical meta-analyses of treatment A vs treatment B, funnel plots as proposed by Sterne and Egger are useful methods for detecting publication bias as they are centred around a treatment difference. In this study, there is no treatment difference, and whilst it is possible to present a funnel plot which is centred around the fixed effect estimate from the metanalysis, it is less meaningful than would be observed in clinical trials data. Therefore, we believe that simply presenting the data and viewing the distribution is probably the best method of detecting publication bias.

Discussion

line 219 (and results lines 204-217)

On average patients can expect a large and sustained improvement in PROMS, but more interesting which percentage has a score equal to or less than the baseline PROM score. Could the authors calculate this from data collected?

This would indeed be an interesting thing to discuss, unfortunately however as no study reported the individual participant PROMs scores, only the group means and SD, we would be unable to calculate this.

In summary the article gives a good overview of to be expected revision and to be expected outcome for patients, addressing the above points will improve the usability of findings in daily practice

We thank you for your kind comment.

Reviewer #4: Thank you for the opportunity to review this systematic review and meta-analysis on the survivorship of total shoulder arthroplasty. The authors discuss a very interesting study, especially as the rates of total shoulder arthroplasty continue to increase. The authors evaluate a number of different cases, from the literature and registry data, helping substantiate their findings. Although the authors provide a great insight to TSA, there are still some edits that might benefit this manuscript:

Thank you, we have made every effort to incorporate your suggested edits.

- 1) Did the authors note any influence of geography on outcomes?

The included studies were from a limited number of geographic locations (see table 1). Therefore, we feel any sub-group analysis would not be useful due to the subsequent reduction in numbers and the risk of bias.

2) Were certain year ranges more associated with longer survivorship (i.e., given all of the advances with TSA over the past few years, it would be expected that more recently performed procedures would fare better than older procedures).

Thank you, this is indeed an interesting question that we discuss earlier in this document in response to another reviewer. Unfortunately, due to the fact we required 10-year mean/median follow-up, many of the recent advances you describe would not be captured in our data set. Once registries further mature, we would hope that this analysis could be undertaken. We did not look for any time trends for survival in the included papers as the overall numbers were small to start with and any sub-group analysis would reduce this further and likely be inaccurate or not generalisable.

3) The authors should clarify the inclusion data range.

Thank you for the suggestion, which is in keeping with one made by another reviewer. We have now included more information on the inclusion dates.

“Six unique series, published between 1998 – 2015, reported survival of 529 total shoulder replacements (TSR) at 13 time points with follow-up ranging from 10 to 21 years (Appendix 2).”

PAGE 9 / PARAGRAPH 3 / LINES 200 - 201

“Four unique series, published between 1998 – 2017, reported survival of 364 shoulder humeral hemiarthroplasties (HAs) at 10 time points with follow-up ranging from 10 to 21 years (Appendix 2).”

PAGE 10 / PARAGRAPH 1 / LINES 209 - 210

4) To help the readership, a more directed "aims"-type statement should be included in the Introduction. This will help keep the manuscript organized.

Thank you for this suggestion. We have amended the manuscript to achieve this.

“We aimed to provide the best quality pooled estimates of implant survival at a minimum 10 years’ follow-up. The decision to revise a poorly performing shoulder replacement is multifactorial that may be sensitive to both patient and surgeon preferences. Therefore, we also aimed to make a pooled estimate of the likely patient reported outcome at long-term follow-up, in essence to answer the question: Will my shoulder be better 10 years after surgery?”

5) What patient reported outcomes measures were utilized?

Thank you, we agree that we had not made this clear enough and now specify explicitly in the manuscript. We have given examples in the methods of the inclusion types for PROMs, and in the results.

“Data extracted were: publication date, baseline population demographics, number of patients (n), surgical indication proportion (% OA and/or % RCTA), follow-up duration (>10 years), implant name and construct type (TSA, HA or RTSA), loss to follow-up, survival estimates (including CIs) and all available Patient Reported Outcome Measure (PROM) (e.g. Visual Analogue Scales (VAS), Constant score, Disabilities of the Arm, Shoulder and Hand (DASH)), data (outcome measure used, (outcome measure used baseline mean score (SD), follow-up duration in 5 year increments, follow-up mean score (SD)).”

“All reported the outcome of shoulder-specific PROMs, without the addition of generic quality of life measures. Five studies reported the Constant score, one the simple shoulder test (SST) and one a four-point linear pain scale previously described by Neer.”

6) Can the authors comment/hypothesize why they think we see these trends in failures? What recommendations can the authors suggest? These answers should help drive meaningful clinical changes orthopaedic surgeons can make in their practices.

As a systematic review of long-term survival, we were really aiming to provide information to answer one of the most commonly asked questions in the orthopaedic clinic, “how long will it last?”. We would be very wary of suggesting a change in practice from our analysis as these are historic case-series. What we feel to be important in our report is that the case-series survival mirrors the registry data which we often presume is more robust, and this is encouraging. Furthermore, we feel that the survival of shoulder implants reported in this paper will be reassuring to surgeons, patients and commissioners of healthcare services. Therefore, future implant and technical developments should continue to be monitored with well performed case-series analyses.

7) From the registry data, how were anatomic total shoulder arthroplasty and reverse total shoulder arthroplasty distinguished. These procedures can have the same CPT codes. What system was used to distinguish these cases?

Thank you for highlighting this. As we did not perform the individual patient level analyses, we're reliant on the accuracy of the reporting source. The Australian registry collect implant level data and group this in their report by class (TSA, RTSA etc).

8) Did the authors note any particular patient populations that had better or worse outcomes?

Thank you for asking this important question. Following a similar query from another reviewer, we have added a further discussion point on age range, notably that the young HA population in case series have survival characteristics that may not be generalisable. They demonstrate lower survival presumably as they are higher functioning.

"It is notable that the demographic characteristics from the case-series and registry data are similar for the TSR group, and concordantly their survival rates are also comparable. For the HA group, the case-series data contain a more male dominated and younger population. All but one of the case-series report an average age of <60yrs, therefore the survival findings from case-series may lack generalisability."

PAGE 13 / PARAGRAPH 1 / LINES 270 - 274

9) The Conclusion could be tightened. There should be a message to physicians and providers on how to use this data for practical purposes.

Thank you for this suggestion. We feel it is important to report our findings as simply as possible to keep with the aims of the paper and avoid overstating conclusions.

"By pooling survival from case-series and registry data, we have been able to provide a reliable estimate of 10-year survival rate of shoulder replacements. We found that over 90% of shoulder replacements last for at least 10 years. Patients experienced sustained and marked benefit to 10 years. This information should be reassuring for patients, health professionals and commissioners of health services."

PAGE 15 / PARAGRAPH 3 / LINES 336 - 339

10) What specific PROMs were evaluated?

As we discuss above three different PROMs were evaluated in the case series. The Constant score, the Simple Shoulder Test (SST) and a linear pain scale as described by Neer.

“All reported the outcome of shoulder-specific PROMs, without the addition of generic quality of life measures. Five studies reported the Constant score, one the simple shoulder test (SST) and one a four-point linear pain scale previously described by Neer.”

PAGE 10 / PARAGRAPH 4 / LINES 231 -232

11) The authors mention results at 15 and 20 years in their Methods and Results, but not so much in the Discussion/Conclusion. These data should also be discussed as results also show to potential for TSA.

Thank you for this comment, it is of course our intention to present and discuss all useful findings to the reader, however, we are very cognisant of the fact that our data is strongest at the 10-year mark and very limited at 20 years, so did not want to overstate findings that may not be as reliable comparatively. We discuss the 20-year data in our case-series discussion. With broad confidence intervals at both the 15- and 20-year analysis, we do not feel that the discussion over TSA superiority can be made.

“This study found very limited extended case-series 20-year data, all from the Mayo group, with survival for TSRs of 83·2% and HAs 75·6%, which are lower than the HES report of 87·8% (95% CI 87·2, 88·4) at 18 years but comparable to the full Mayo Clinic registry of 81·4% (95% CI 78·4, 84·5) for TSR, but worse than the HA survival of 85·0% (95% CI 81·8, 88·4) at a 20 years, notably there is a younger age cohort in their HA case-series.”

PAGE 13 / PARAGRAPH 1 / LINES 265 - 270