Predictors of clinical outcome in total hip and knee replacement : a methodological appraisal of implants and patient factors

Keurentjes, J.C.

Predictors of Clinical Outcome in Total Hip and Knee Replacement

A METHODOLOGICAL APPRAISAL OF IMPLANTS AND PATIENT FACTORS

J. Christiaan Keurentjes

PhD thesis, Leiden University Medical Center, Leiden, The Netherlands.

ISBN: 978-90-5335-908-2.

Cover: "Study for Arithmetic Composition", Theo van Doesburg - www.geheugenvannederland.nl
Licensed under Public domain via Wikimedia Commons.

Layout: J. Christiaan Keurentjes.

Printed by: Ridderprint BV.

The research presented in this thesis was funded by an ongoing research grant of the Dutch Arthritis Foundation to Prof. Dr. Nelissen, Department of Orthopaedic Surgery, Leiden University Medical Center, grant number LLP-013.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscripts.

The printing of this publication is financially supported by the Dutch Arthritis Foundation, Nederlandse Orthopaedische Vereniging, Annafonds NOREF, Chipsoft, Hanssen Footcare.

Predictors of Clinical Outcome in Total Hip and Knee Replacement

A METHODOLOGICAL APPRAISAL OF IMPLANTS AND PATIENT FACTORS

Proefschrift

ter verkrijging van de graad van Doctor
aan de Universiteit van Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op
dinsdag 30 september 2014
klokke 16:15 uur

door

Johan Christiaan Keurentjes

geboren te Satu Mare, Roemenië
in 1982

Promotiecommissie

Promotor:	Prof. dr. R.G.H.H. Nelissen
Copromotor:	Dr. M. Fiocco
Overige leden:	Prof. dr. S.K. Bulstra (UMCG, Groningen)
	Prof. dr. H.C.W. De Vet (VUMC, Amsterdam)
	Dr. A. Sedrakyan (Weill Cornell Medical College, New York)
	Prof. dr. F.W. Dekker
	Prof. dr. T.P.M. Vliet Vlieland
	Dr. P.J. Marang - Van De Mheen

Contents

1

Introduction

Total Hip Replacement (THR) and Total Knee Replacement (TKR) are effective surgical interventions, which alleviate pain and improve Health-Related Quality of Life (HRQoL) in patients with hip or knee joint degeneration.[1] National joint replacement registries show good long-term results regarding the probability of revision surgery, which is lower than 10% at 10 years follow-up, both for THR and TKR.[2, 3] Compared to non-operative treatment, both THR and TKR have been shown to be cost-effective interventions.[4–7]

Epidemiology Annually, 25,000 THR and 20,000 TKR are performed in the Netherlands.[8] These numbers are projected to rise substantially, due to demographical changes, the rising incidence of overweight and obesity, improved long-term outcomes of joint replacements, more active lifestyle of the elderly and the increasing number of orthopaedic surgeons.[9] The annual numbers of THR and TKR performed are expected to increase to approximately 50,000 and 60,000 in 2030.[9]

In the Netherlands, the majority (80%) of THR are performed for osteoarthritis (OA).[8] Less frequent indications for joint replacement include a displaced femoral neck fracture, osteonecrosis, secondary posttraumatic OA, rheumatoid arthritis, OA due to Legg-Calvé-Perthes disease and the treatment of a neoplasm.[8] The mean age at joint replacement is 70 years, two-thirds of the patients are females.[8] The majority (96%) of TKR are also performed for OA.[8] Less frequent indications for joint replacement include rheumatoid arthritis, secondary posttraumatic OA, osteonecrosis and the treatment of a

neoplasm.[8] The mean age at joint replacement is 68 years, two-thirds of the patients are females.[8]

Implants Currently, a wide variety of Total Hip Implants is available to orthopaedic surgeons worldwide. The probability of revision surgery varies considerably between different implants.[10] In order to prevent unnecessary harm and limit secondary health care costs, it is imperative to choose an implant with a low probability of revision surgery, when performing primary THR.

Patient and Surgeon Factors Although joint replacements are highly effective in improving HRQoL and joint specific functioning at the group level,[1] this is not the case for each individual patient. Persistent pain is reported in 9% of THR patients and 20% of TKR patients at long term follow-up.[11] Additionally, up to 30% of patients are dissatisfied with the results after surgery.[12–20]

The therapeutic options for patients with an unfavourable outcome after THR or TKR are limited. The outcome of revision surgery performed without a specific mechanical or physiological indication is highly unpredictable. Predicting which patient groups are at increased risk of an unfavourable outcome after joint replacement may provide additional insights in the mechanisms involved and offer the possibility of intervention in order to optimise the outcome.[21] At the very least, it allows patients to be well informed of their specific risks and expected gains before surgery.

In this thesis, we have investigated the role of two potential predictors, which are inexpensive to measure and easily available in clinical practice. Firstly, we studied whether the patients' Socio-Economic Position was associated with the improvement in Health-Related Quality of Life and patient satisfaction after THR and TKR. Secondly, we studied whether the preoperative radiographic severity of OA was associated with the improvement in Health-Related Quality of Life and patient satisfaction after THR and TKR.

Research Methodology

Competing Risks: The probability of revision surgery at a specific point in time (given

that revision surgery has not occurred up to that point in time) is of special interest. The Kaplan-Meier estimator is often used to estimate this probability.[22] This method assumes independence of the time to event and the censoring distribution. In the presence of competing events, this assumption is violated.

Clinimetrics: Health-Related Quality of Life can be studied at multiple levels. Minimal Clinically Important Differences (MCIDs), Clinically Important Differences (CIDs) and Patient Acceptable Symptom States (PASS) are closely related concepts, which could provide more insight into the patients outcome at the individual level. MCIDs are defined as the *minimal* improvement in a specific outcome measure, which is perceived by patients as beneficial or harmful. The CID constitutes a larger, more clinically relevant improvement. In PASS, the focus is shifted from the improvement to the actual outcome achieved.

Questionnaire Mode Preference: Electronic forms of data collection have gained interest in recent years.[23] Expected advantages include more complete data capturing, immediate availability of results and less costs in administrating and entering data.[23, 24] However, electronic questionnaires might induce selection bias, as some patients could be less inclined to participate in a study which exclusively uses electronic questionnaires.

The Paprika Study In order to study predictors of clinical outcome in THR and TKR, we set up the Paprika Study: "Patients Prospectively Recruited in Knee and Hip Arthroplasty" (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189; Netherlands Trial Register: NTR2190). Patients who previously participated in the Trigger Study or the TOMaat Study, both multicenter randomised controlled trials, were eligible for inclusion in the Paprika Study. The Trigger Study compared the effect of a restrictive blood transfusion policy compared to standard care on the red blood cell transfusion rate after THR and TKR.[25, 26] The TOMaat Study compared the effect of different blood management modalities on the red blood cell transfusion rate during and after THR and TKR (Netherlands Trial Register: NTR303). Patients who were willing to participate in the Paprika Study, were sent a questionnaire and a saliva DNA collection kit.

In this thesis, we focussed on the improvement in Health-Related Quality of Life after primary THR and TKR. Therefore, only patients who previously participated in the TOMaat Study were available for analyses, as Health-Related Quality of Life was not measured before joint replacement in the Trigger Study. A comprehensive overview of the study population of each paper, in which data from the Paprika Study was used, is presented in figure 1.1 on the facing page.

Thesis Overview In chapter 2 (p. 9), we have systematically searched and appraised the literature to compare the probability of revision surgery at 10 years follow-up for each THI to the National Institute for Clinical Excellence (NICE) benchmarks. Based on this systematic review of the literature, we can recommend a number of THI for primary THR, which outperform NICE benchmarks.

Using data from the Paprika Study, we investigated whether the patients Socio-Economic Position affects the improvement in HRQoL and satisfaction with the surgical results in chapter 3 (p. 37). In chapter 4 (p. 55), we investigated whether the preoperative radiographic severity of OA affects the improvement in HRQoL and satisfaction with the surgical results, using data from the Paprika Study.

In chapter 5 (p. 71), we assessed how much bias is introduced by the Kaplan-Meier estimator in a long-term cohort study.

In chapter 6 (p. 81), we performed a systematic review to find studies reporting MCIDs in HRQoL after primary or revision THR and TKR. In chapter 7 (p. 93), we determined CIDs in HRQoL after primary THR and TKR, using data from the Paprika Study. In chapter 8 (p. 105), we determined PASS in joint specific functioning scores after primary THR and TKR, using data from the Paprika Study.

In chapter 9 (p. 119), we assessed which questionnaire mode THR and TKR patients preferred in the Paprika Study.

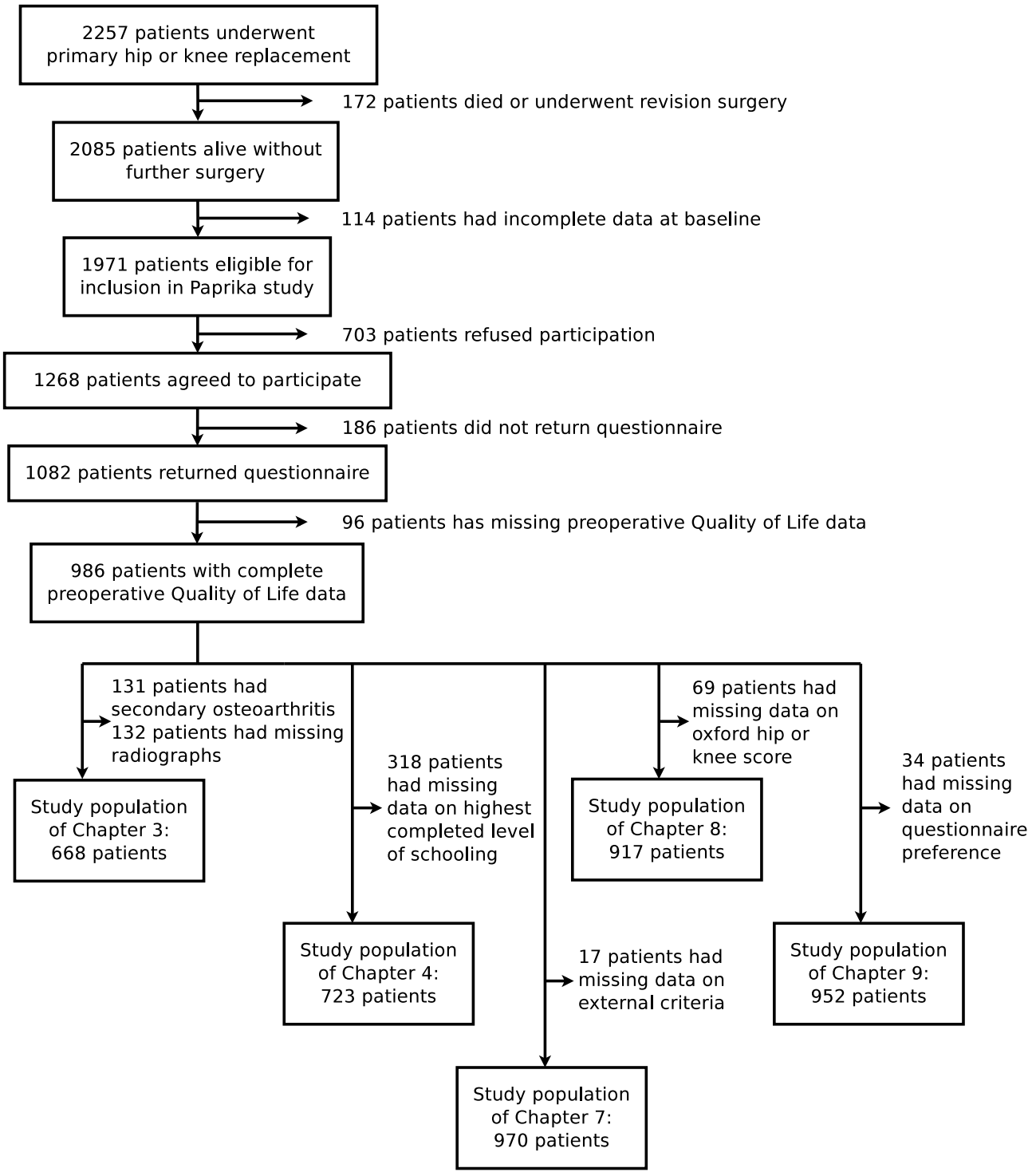


Figure 1.1: **Flow-chart of Paprika Study.** Study population of Chapter 3, 4, and 7 – 9.

Part I

Implants

2

Which Implant should we use for Primary Total Hip Replacement?

JC Keurentjes[1], BG Pijls[1], FR Van Tol[1], JF Mentink[1], SD Mes[1], JW Schoones[2], M Fiocco[3], A Sedrakyan[4], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Walaeus Library, Leiden University Medical Center.

3 Medical Statistics and BioInformatics, Leiden University Medical Center.

4 Public Health, Weill Cornell Medical College, New York City.

Accepted in Journal of Bone and Joint Surgery.

Abstract

Background Many total hip implants are currently available on the market worldwide. We aimed to estimate the probability of revision surgery at ten years for each individual total hip implant and to compare these estimates with the National Institute for Health and Care Excellence (NICE) benchmark.

Methods We performed a meta-analysis of cohort studies. The methodological quality was assessed with use of the Assessment of Quality in Lower Limb Arthroplasty (AQUILA) checklist. We searched PubMed, EMBASE, Web of Science and the Cochrane Library. Additionally, National Joint Registries that were full members of the International Society of Arthroplasty Registers (ISAR) were hand searched. Studies in which the authors reported the survival probability for either the acetabular or the femoral component of primary total hip replacements, with at least 100 implants at baseline, and in which at least 60% of the patients had primary osteoarthritis were eligible for inclusion.

Results The search strategy revealed 5513 papers describing survival probabilities for thirty-four types of acetabular components and thirty-two types of femoral components. Eight types of acetabular cups and fifteen types of femoral stems performed better than the NICE benchmark.

Conclusions We recommend the surgeons performing a primary total hip replacement use an implant that outperforms the NICE benchmarks.

Introduction

Total Hip Replacement (THR) is an effective surgical intervention to alleviate pain, restore functionality of the hip and improve the quality of life of patients with end-stage degeneration of the hip joint.[1, 21, 27, 28]. Currently, a wide variety of Total Hip Implants (THI) is available to orthopaedic surgeons worldwide.[29] Many factors, such as the cost of the implant, familiarity with the design and instruments and ease of use, influence the choice for a particular THI. Arguably, from both a patient and a societal perspective, the most important factor is the clinical performance of the total hip implant

and the probability of revision surgery during a given period of time. Revision hip arthroplasty is technically challenging with a higher complication rate, a longer hospital stay, and a higher cost than primary total hip replacement and can lead to disability and death.[30–34] Clearly, choosing a total hip implant that is associated with the lowest rate of revision surgery can prevent harm and reduce long-term health-care costs. Recently, the National Institute for Health and Care Excellence (NICE) suggested a ten-year revision rate of $\leq$10% as an acceptable benchmark performance of a primary total hip implant, which was loosely based on an earlier report by Murray et al.[29, 35]

The objective of our study was to systematically search and appraise the literature to estimate the probability of revision surgery at ten years for each individual type of total hip implant. Additionally, we sought to compare the estimates of the probability of revision surgery for each total hip implant to with NICE benchmark.

Materials and Methods

Protocol and registration This systematic review and meta-analysis was performed from March 2011 to February 2013, with use of the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement for development of the study protocol and reporting the results of our study.[36]

Eligibility criteria The NICE Technology Appraisal Guidance states: "The evidence used in support of any prosthesis . . . should relate to data on 10 or more years follow up from a number of centers, obtained via adequately sized, well conducted observational studies (preferably with consecutive patients from non-selected populations) or randomised controlled trials. Such evidence should have been published or be available for peer review".[35] Studies in which the authors reported the survival probability (i.e. the Kaplan-Meier estimate) for either the acetabular or the femoral component of a primary total hip replacement with use of revision for any reason or for aseptic loosening at ten years as the end point were eligible for inclusion. We considered studies to be of adequate size when there were at least 100 implants at baseline, and we defined a study

population as representative of the general population at large when at least 60% of the patients had primary osteoarthritis. Studies with fewer than 100 implants at baseline and in which $<$60% of the patients had primary osteoarthritis were excluded. Studies were also excluded when the authors described the outcomes of multiple (sub)types of implants without reporting the outcomes for each (sub)type separately. Articles written in any language other than English, Dutch, German, French, Spanish or Italian were not eligible for inclusion. In order to limit the extent of publication bias, no publication status restrictions were imposed.

Information sources On March 22, 2011, an experienced independent information specialist (JWS) searched four electronic databases: PubMed, EMBASE, Web of Science and the Cochrane Library. We also performed implant-specific PubMed searches for all primary total hip implants registered in the first annual report of the Dutch Arthroplasty Register (Landelijke Registratie Orthopaedische Implantaten [LROI]).[8] Finally, National Joint Registries that were full members of the International Society of Arthroplasty Registers (ISAR, www.isarhome.org) were hand searched.

Search strategy The following search terms were applied to Pubmed and adapted for all other databases:
(tha[tw] OR “total hip” OR “total hips” OR ((“total joint” OR “total joints”) AND (hips OR hip)) OR (total hip AND (prosthesis OR prosthetic OR endoprosthesis OR endoprostheses OR endoprosthetic OR arthroplasty OR arthroplasties OR replacement [tiab])) OR (Hip Replacement Arthroplasty AND total [tiab]) OR Hip Replacement Arthroplasty OR hip arthroplasty OR hip replacement OR Hip Prosthesis) AND (Osteoarthritis OR Osteoarthritides OR osteoarthriti* OR Osteoarthrosis[tiab] OR Osteoartroses OR athrosis[tw] OR arthroses OR “Degenerative Arthritis”) NOT (early[tw] OR initial[tw] OR preliminary[tw] OR “short follow-up”[tw] OR “Letter”[Publication type] OR “Case Reports”[Publication Type])

Study Selection Two authors (JFM & SDM) independently screened the titles and abstracts of the search results using pre-specified eligibility criteria, as stated above.

Two other authors (JCK & FRvT) screened the full text of the remaining articles using the same eligibility criteria. Disagreements between authors were resolved by consensus.

Data collection process and data items Data collection was performed by two authors (JCK & FRvT) independently using predefined data extraction sheets. Inconsistencies between the two authors were resolved by consensus. When data were not reported numerically, but were presented graphically in Kaplan-Meier curves, the estimated observations of both authors were averaged. The brand name and manufacturer of the implant, the Kaplan-Meier estimate at ten years, and its standard error and 95% confidence interval were extracted from each included study.

Risk of bias in individual studies The methodological quality of all included studies was assessed using the Assessment of Quality in Lower Limb Arthroplasty (AQUILA) checklist, a tool specifically designed to appraise the quality of observational studies concerning total hip replacement and total knee replacement.[37] Two authors (JCK & FrvT) independently assessed the quality of all included studies, using predefined data extraction sheets. Inconsistencies between the two authors were resolved by consensus.

Summary measures and Synthesis of results The principal summary measure was the survival probability for each implant at 10 years with use of revision for any reason as the end point. The secondary summary measure was the survival probability for each implant at 10 years with use of revision for aseptic loosening as the end point. Estimates of the survival probability in different studies on the same implant were pooled with use of inverse variance weighting. When no estimate of the variance or standard error of the survival probability at 10 years was presented, we deduced the missing standard error from the confidence interval of the survival probability. When the study did not provide an estimate of the variance or standard error, or a confidence interval, we imputed the missing standard error from the mean standard error of all other studies.[38, 39] When $>$50% of all standard errors were missing, we imputed the missing standard errors with single imputation on the basis of the survival estimate and the number of

implants at baseline. We chose this approach instead of a more elaborate modelling approach[40–42] for two reasons. First, we were interested in the survival probability at only one specific point in time. Second, the majority of studies that did not provide the standard error also did not give enough information to allow modelling of the survival probability.

In order to test whether each implant performed better than the NICE benchmark, we calculated the confidence interval for each implant survival estimate. The 10-year revision rate of 10% for a total hip implant corresponds with a survival probability of 90% for a THI. Therefore, the survival probability of a cup or stem should exceed 90%. Assuming independence of the survival probability for either the cup or the stem, we can summarise the minimal survival probability with the formula: $p_{cup} * p_{stem} \geq 0.9$. When it is assumed that $p_{cup} = p_{stem}$, then the minimal survival probability for the cup is $p_{cup}^2 = 0.9$, leading to a minimal cup survival probability of $\sqrt{0.9}$, which is rounded to 95%. Therefore, the survival probability of either the cup or the stem should exceed 95%. When both the survival estimate and the lower limit of the confidence interval were $>$95%, we concluded that that particular implant performed significantly better than the NICE benchmark. When both the survival estimate and the upper limit of the confidence interval were $<$95%, we concluded that that particular implant performed significantly worse than the NICE benchmark. In all other cases, we concluded that there was insufficient evidence to suggest that the particular implant performs better or worse than the NICE benchmark.

All analyses were performed with use of R, version 2.15.2.[43]

Source of funding This study was funded by a grant from the Dutch Arthritis Association (Grant number LLP-13), which had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all of the data in the study. All authors had final responsibility for the decision to submit for publication.

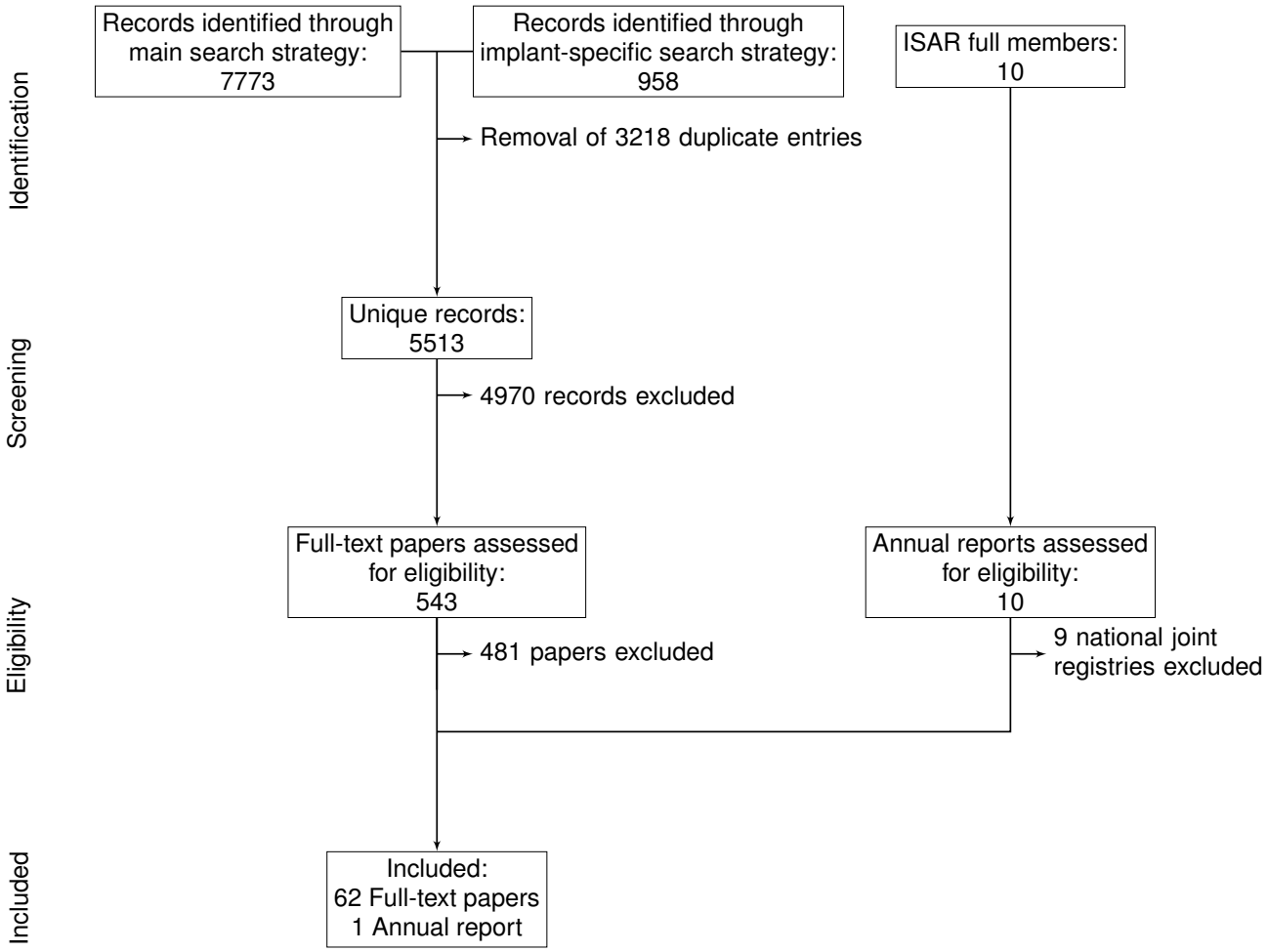


Figure 2.1: **Flow-chart of study inclusion.**

Results

Study selection Our search strategy revealed 8731 hits: the main search strategy yielded 7773 hits and the implant-specific search yielded an additional 958 hits (Fig. 2.1). After removal of duplicate entries, 5513 unique papers remained, and their titles and abstracts were screened. 4970 papers were excluded, leaving 543 papers eligible for inclusion. Further assessment of eligibility based on the full-text papers led to the exclusion of 481 papers: 259 papers did not report the survival probability of a cup or stem at ten years, and 222 papers did not provide separate results for cups or stems. Of all national joint registries that were full members of ISAR, only the Swedish Hip Arthroplasty Register reported separate results for cups and stems at ten years with revision for any reason as the end point.[44] This left 63 papers for further analysis.

Study characteristics and Risk of bias Tables 2.1 (p. 22) and 2.2 (p. 23) provide an overview of the characteristics of all included studies in which the end point was revision for any reason, and tables 2.3 (p. 24) and 2.4 (p. 25) provide such an overview for the studies in which the end point was revision for aseptic loosening. The methodological quality of the studies in which revision for any reason was the end point is shown in tables 2.5 (p. 26) and 2.6 (p. 27), and the methodological quality of the studies in which the end point was revision for aseptic loosening is shown in tables 2.7 (p. 28) and 2.8 (p. 29). We found 41 cohorts for which the ten-year survival probability of an acetabular cup was described with revision for any reason as the end point. These studies included a total of 34 different acetabular implants: ABG 1,[45, 46] ABG 2 HA,[44] ACS Triloc+,[47] Arthopor,[47] ATLAS II,[48] ATLAS III,[48] Charnley,[44, 49] Charnley Elite,[44] CLS Spotorno,[44] Conserve Plus,[50] Exeter All-Polyethylene,[51] Exeter Duration,[44] Fitmore,[52] Harris Design 2,[53] Harris-Galante I,[54] Harris-Galante II,[45] Hofer-Imhof,[55] JRI Threaded Cup,[56] Lubinus,[44] Lubinus Eccentric,[57] Mallory-Head Cementless,[45, 58] Miami Orthopaedic Surgical Clinic: All Polyethylene,[49] Miami Orthopaedic Surgical Clinic: Metal Backed,[49] Morscher Press Fit,[59, 60] Mueller,[49] Novae,[61] PCA Pegged,[45] Plasmacup,[62] RM,[63] Romanus,[45] T28,[49, 64] Trilogy HA,[44] Universal,[45, 65] Zweymuller-Alloclassic Screw Cup.[66, 67]

42 papers described the ten-year survival probability of 32 different femoral stem implants with revision for any reason as the end point: ABG 1,[44–46] Anatomic Mesh,[45] Bicontact,[68] Bi-Metric,[45, 69, 70] Charnley,[44, 71] CLS Spotorno,[44, 45] Conserve Plus,[50] Corail,[72] Exeter Polished,[44] Exeter Universal,[73] Freeman Cementless,[74] Furlong,[56] Harvard Femoral Stem,[75] Heritage,[76] Iowa polished,[76] Lord Madreporique,[45, 77] Lubinus IP,[57, 78] Lubinus SP,[57, 78] Lubinus SPII,[44] Mallory Head Cemented,[79] Mallory Head Cementless,[58, 80] MS-30,[59] Omnifit Cemented,[71] Osteonics Cementless,[81] PCA,[45] Profile Porous,[45] R-B Interlok,[71] SBG,[82] Stanmore Custom Made,[83] Taperloc,[84] Triumph,[76] Zweymuller SL.[66]

39 papers described the ten-year survival probability of 31 different acetabular cup implants with revision for aseptic loosening as the end point: ABG 1,[45] ACS Triloc+,[47] Arthopor,[47] Brunswik,[85] Charnley,[49, 85–87] Charnley All-Polyethylene,[88] Conserve Plus,[50] Elite Ogee,[73] Exeter,[87] Exeter All-Polyethylene,[51] Fitmore,[52] Harris Design-2 All-Polyethylene,[89] Harris-Galante II,[45] Hofer-Imhof,[55, 90] JRI Threaded Cup,[56] Link,[87] Lord Threaded,[77] Mallory-Head Cementless,[45] Miami Orthopaedic Surgical Clinic: All Polyethylene,[49] Miami Orthopaedic Surgical Clinic: Metal Backed,[49] Morscher Press Fit,[59, 60] Mueller,[49] PCA Pegged,[45] Romanus,[45] Spectron,[87] T-28,[49, 64] Titan,[87] Trabecular Metal Monoblock Acetabular Component System,[91] Universal,[45, 65] Weber Hemispheric,[92] Zweymuller-Alloclassic Screw Cup.[93, 94]

Finally, we found 52 cohorts in which the ten-year survival probability of 37 different femoral stem implants was assessed with revision for aseptic loosening as the end point: ABG 1,[45, 46] Anatomic Mesh,[45] Bi-Metric,[45, 69, 70] Charnley,[71, 87, 95] Charnley Elite-Plus,[96, 97] Charnley Flat-back,[97] CLS Spotorno,[45] Corail,[72] Exeter,[87] Exeter Matt,[97] Exeter Universal,[73, 97, 98] Freeman Cemented,[99, 100] Freeman Cementless,[74, 101] Furlong,[56] Harris Design 2,[89] Harvard Femoral Stem,[75] Interlok,[97] ITH,[87] Lord Madreporique,[45, 77] Lubinus IP,[97] Lubinus SP II,[97] Mallory Head Cementless,[58, 80] MS-30,[59] Muller Straight Protasul-10 Cobalt-Nickel-Chromium,[97, 102, 103] Muller Straight Protasul-100 Titanium,[104] Muller Style Titanium,[105] Omnifit Cemented,[71] Osteonics Cemented,[106] Osteonics Cementless,[81] PCA,[45] Profile Porous,[45] R-B Interlok,[71] SBG,[82] Stanmore Custom Made,[83] Taperloc,[65, 84] Titan,[87] Zweymuller Alloclassic.[94]

Synthesis of results An overview of the survival probability for the different implants is presented in figures 2 through 5. With use of revision for any reason as the end point, the following acetabular cups performed better than the NICE benchmark: JRI Threaded Cup, Conserve Plus, Zweymuller-Alloclassic Screw Cup, Charnley Elite, Lubinus, Exeter Duration, Charnley, T28 (fig. 2.2 (p. 30)). With use of revision for any reason as the end point, the following femoral stems performed better than the NICE benchmark:

Stanmore Custom Made, MS-30, Iowa Polished, RB Interlok, Taperloc, Corail, Furlong, SBG, Zweymuller SL, CLS Spotorno, Mallory-Head Cementless, Osteonics Cementless, Lubinus SPII, ABG 1, Exeter Polished (fig. 2.3 (p. 31)). With use of revision for aseptic loosening as the end point, the following acetabular cups performed better than the NICE benchmark: Weber Hemispheric, Trabecular Metal Monoblock Acetabular Component System, JRI Threaded Cup, Fitmore, Conserve Plus, Morscher Press Fit, Zweymuller-Alloclassic Screw Cup, Arthropor, ACS Trilok+, Charnley, Titan, Spectron, Charnley All-polyethylene, Exeter (fig. 2.4 (p. 32)). With use of revision for aseptic loosening as the end point, the following femoral stems performed better than the NICE benchmark: ABG 1, Osteonics Cementless, RB Interlok, Zweymuller Alloclassic, Freeman Cementless, Stanmore Custom Made, MS-30, Corail, Profile Porous, Bimetric, Mallory-Head Cementless, Taperloc, Omnifit Cemented, Furlong, CLS Spotorno, Harris Design II, Exeter, Exeter Universal, Titan, Osteonics Cemented, Freeman Cemented, Charnley, Muller Straight Protasul-10 Cobalt-Nickle-Chromium (fig. 2.5 (p. 33)).

Discussion

In this systematic review and meta-analysis, we estimated the probability of revision surgery at ten years for 34 types of acetabular cups and 32 types of femoral stems that were available on the market with published results. Of these implants, 8 acetabular cups and 15 femoral stems performed better than the NICE benchmark.
Most studies were of low methodological quality: the majority of studies consisted of non-consecutive cohorts, with more than 5% of the hips lost to follow-up and no worst-case analysis.

In the past decades, numerous efforts have been made to improve the survival probability of primary total hip implants. Some efforts, such as the addition of antibiotics to bone cement,[107] have led to an improvement in survival probability. Others, such as the introduction of Boneloc® cement or the 3M Capital Hip System have led to unprecedented failures, which could have been prevented by phased introduction of new implants and techniques.[108, 109] Despite subsequent calls for stricter regulation

of new total hip implants,[29, 110] few actions were taken, facilitating the recent disaster with the ASR hip prosthesis.[111] In providing an overview of all implants that perform better or worse than the NICE benchmark, we aid practising orthopaedic surgeons in choosing safe, time-proven implants for primary total hip replacement. Additionally, our study documents that an astonishing limited number of publications are currently available.

There are a number of limitations to this study. The description of the type of implant used in a specific total hip replacement cohort was often limited to the specific brand name. Some studies, such as the well described one by Franklin et al,[112] included the results of multiple subtypes of implants, which had undergone major changes in design, summarized in one survival estimate at 10 years. We excluded studies which did not specify separate survival estimates for subtypes with substantial changes in design. In all other papers, we assumed that a single subtype of implant was used for all patients. Additionally, the NICE benchmarks were poorly defined, leaving much room for interpretation.[35] The recommendation to use a ten-year revision rate of 10% as a threshold does not specify a particular end point — e.g. aseptic loosening or revision for any reason. Additionally, it is unclear whether this revision rate should include the acetabular or femoral component or both implants. Furthermore, no guidance is given regarding the statistical methods to use for outlier detection. Finally, the guidelines do not define “adequately sized” or “well conducted” studies. In order to err on the side of caution, we chose revision for any reason instead of revision for aseptic loosening as the primary outcome measure. Aseptic loosening is considered the principal mechanism of failure at the time of long-term follow-up, is slowly progressive and causes disabling pain.[113] Especially in the case of focal osteolysis, an implant might appear solidly fixed at revision surgery, despite moving up to 1 mm relative to the surrounding bone.[114] In order to minimise the risk of misclassification bias (e.g. misclassifying cases of aseptic loosening as cases revised for persistent pain after joint replacement), we chose to use revision for any reason as the principal outcome measure.[115] On the basis of our clinical judgement, we defined the revision rate of 10% as referring to the combination of

both implants and defined “adequately sized” as a minimum of 100 implants at baseline. No competing risk analyses were performed in any of the included cohorts. It is highly unlikely that no competing events, such as the death of a patient, have taken place within 10 years after primary THR. Disregarding these competing events leads to an underestimation of the survival probability.[116] Therefore, some implants might outperform the NICE benchmark in reality but not appear to do so on the basis of their survival estimates because of unrealistic statistical assumptions.

In our analyses, we assumed that the case mix of all studies was similar. Regarding one of the most important characteristics— namely, the indication for joint replacement— this was certainly the case, as this was one of the inclusion criteria. Other characteristics, such as age, sex, physical activity, and number of co-morbidities were not recorded and might have differed among the cohorts. Some patient characteristics, such as age and sex, are easily identified in most studies. Others, such as physical activity and number of co-morbidities are not uniformly measured if they are measured at all. Because there is no current consensus on relevant case-mix variables,[37] we decided to omit these variables from this systematic review.

Finally, the majority of the survival estimates were based on a single study, performed in a single center. This raises the question of whether the survival rates presented in this meta-analysis represent the actual survival rates of these implants. In the unlikely case of extreme publication bias, an implant might appear to outperform the NICE benchmark in the only published study, while performing worse in the unpublished reality. Extreme negative publication bias is also theoretically possible. Surgeons who notice poor results using a certain prosthesis might be more inclined to report their results, as a general warning, than are surgeons who notice acceptable results. In the case of extreme negative publication bias, an implant might appear to perform worse than the NICE in the only published study, while performing better in the unpublished reality. It is difficult to study the effect of publication bias in this meta-analysis. Conventional methods such as funnel plots would fail in this case, as it would be pointless to make a funnel plot for an implant for which there was only one estimate and therefore only one

point. A sensible approach is to interpret estimates based on the experience at a single center with more caution, especially if those centers were involved in the design of the implant.[117]

A wide variety of implants is available to orthopaedic surgeons worldwide, but there is a very limited amount of evidence for some of these implants. In the European Union, there is a single organisation for the approval of drugs— the European Medicines Agency, which demands evidence of safety and efficacy in controlled trials. In contrast, for medical devices such as an orthopaedic implant, it is only necessary to obtain a European Conformity (Conformité Européenne (CE)) mark, which requires limited or no evidence of clinical efficacy.[118] Since the introduction of Charnley's total hip replacement in the late 1950s, new successful total hip replacement implants have been designed, lowering the probability of revision surgery. However, recent problems with several hip prostheses have illustrated that patient safety can be at risk when new total hip replacement implants are developed.[119, 120] We encourage the development of new implants but not at the cost of patient safety.[121] Therefore, the development of new implants should take place in the setting of comparative clinical studies. Ideally, results of experimental implants should be compared with results of implants that outperform the NICE benchmark. To provide access to innovative treatments while ensuring evidence is collected, health-care funders need to implement a payment-with-evidence-development approach.[122]

The use of optimally performing total hip implants is possible despite older and more recent disasters with certain hip implants. It is the surgeon who has to decide which implant will provide the best quality for his or her specific patient. The current study underscores that there is evidence in the literature, but that evidence has to be used.

Cup	Ref.	Manufacturer	Country of Study Origin	n at baseline
ABG 1	[45]	Howmedica	Finland	108
ABG 1	[46]	Howmedica	Wales	100
ABG II HA	[44]	n.s.	Sweden	213
ACS Triloc+	[47]	DePuy	USA	394
Arthopor	[47]	Joint Medical Products	USA	433
ATLAS II	[48]	n.s.	France	171
ATLAS III	[48]	n.s.	France	126
Charnley	[49]	Thackrey	USA	238
Charnley	[44]	n.s.	Sweden	23272
Charnley Elite	[44]	n.s.	Sweden	9456
CLS Spotorno	[44]	n.s.	Sweden	1169
Conserve Plus	[50]	Wright Medical Technology	USA	100
Exeter All-Polyethylene	[51]	Stryker	UK	263
Exeter Duration	[44]	n.s.	Sweden	11712
Fitmore	[52]	Sulzer	UK	119
Harris Design 2	[53]	Howmedica	Sweden	126
Harris-Galante I	[54]	Zimmer	Denmark	324
Harris-Galante II	[45]	Zimmer	Finland	277
Hofer-Imhof	[55]	n.s.	Austria	678
JRI Threaded Cup	[56]	JRI	UK	112
Lubinus	[44]	n.s.	Sweden	76047
Lubinus Eccentric	[57]	Waldemar-Link	Finland	444
Mallory-Head Cementless	[45]	Biomet	Finland	110
Mallory-Head Cementless	[58]	Biomet	Canada	307
MOSC[1]: All Polyethylene	[49]	Biomet	USA	100
MOSC[1]: Metal Backed	[49]	Biomet	USA	134
Morscher Press Fit	[59]	Zimmer	Switzerland	124
Morscher Press Fit	[60]	Sulzer	New Zealand	125
Mueller	[49]	Depuy International Ltd	USA	141
Novae	[61]	SERF	France	135
PCA Pegged	[45]	Howmedica	Finland	122
Plasmacup	[62]	B Braun Ltd	UK	318
RM	[63]	Mathys	Netherlands	630
Romanus	[45]	Biomet	Finland	114
T28	[49]	Zimmer	USA	559
T28	[64]	Zimmer	USA	132
Trilogy HA	[44]	n.s.	Sweden	1196
Universal	[45]	Biomet	Finland	898
Universal	[65]	Biomet	USA	114
ZA[2] Screw Cup	[66]	Sulzer	Germany	320
ZA[2] Screw Cup	[67]	Sulzer	Germany	139

Table 2.1: **Study Characteristics of all included Studies, describing the Survival Probability of Acetabular Cups for Revision Any Reason.** n.s.: not specified.
1: Miami Orthopaedic Surgical Clinic. 2: Zweymuller-Alloclassic.

Stem	Ref.	Manufacturer	Country of Study Origin	n at baseline
ABG 1	[45]	Stryker	Finland	390
ABG 1	[46]	Howmedica	UK	100
ABG 1	[44]	n.s.	Sweden	370
Anatomic Mesh	[45]	Zimmer	Finland	135
Bicontact	[68]	B.Braun-Aesculap	Germany	250
Bi-Metric	[69]	Biomet	Sweden	115
Bi-Metric	[45]	Biomet	Finland	1982
Bi-Metric	[70]	Biomet	USA	129
Charnley	[71]	Johnson & Johnson	USA	160
Charnley	[44]	n.s.	Sweden	23272
CLS Spotorno	[45]	Sulzer-medica	Finland	108
CLS Spotorno	[44]	n.s.	Sweden	1169
Conserve Plus	[50]	Wright Medical Technology	USA	100
Corail	[72]	DePuy	France	120
Exeter Polished	[44]	n.s.	Sweden	11712
Exeter Universal	[73]	Howmedica	UK	230
Freeman Cementless	[74]	Finsbury Instruments	UK	100
Furlong	[56]	JRI	UK	134
Harvard Femoral Stem	[75]	Harvard Health Care	UK	269
Heritage	[76]	Zimmer	USA	283
Iowa polished	[76]	Zimmer	USA	120
Lord Madreporique	[45]	Benoist Girard	Finland	286
Lord Madreporique	[77]	Benoist Girard	Norway	116
Lubinus IP	[78]	Waldemar Link	Finland	280
Lubinus IP	[57]	Waldemar Link	Finland	257
Lubinus SP	[78]	Waldemar Link	Finland	263
Lubinus SP	[57]	Waldemar Link	Finland	185
Lubinus SPII	[44]	n.s.	Sweden	76047
Mallory Head Cemented	[79]	Biomet	USA	102
Mallory Head Cementless	[80]	Biomet	USA	2000
Mallory Head Cementless	[58]	Biomet	Canada	307
MS-30	[59]	Zimmer	Switzerland	124
Omnifit Cemented	[71]	Osteonics	USA	305
Osteonics Cementless	[81]	Stryker	USA	226
PCA	[45]	Howmedica	Finland	111
Profile Porous	[45]	Depuy	Finland	115
R-B Interlok	[71]	Biomet	USA	235
SBG	[82]	Plus Orthopaedics	Austria	230
Stanmore Custom Made	[83]	Depuy	Italy	129
Taperloc	[84]	Biomet	USA	129
Triumph	[76]	Zimmer	USA	148
Zweymuller SL	[66]	Zimmer	Germany	320

Table 2.2: **Study Characteristics of all included Studies, describing the Survival Probability of Femoral Stems for Revision Any Reason.** n.s.: not specified.

Cup	Ref.	Manufacturer	Country of Study Origin	n at baseline
ABG 1	[45]	Howmedica	Finland	108
ACS Triloc+	[47]	DePuy	USA	394
Arthopor	[47]	Joint Medical Products	USA	433
Brunswik	[85]	n.s.	Sweden	151
Charnley	[86]	DePuy	Norway	9186
Charnley	[85]	n.s.	Sweden	204
Charnley	[87]	n.s.	Norway	14842
Charnley	[49]	Thackrey	USA	238
Charnley All-Polyethylene	[88]	Zimmer	USA	193
Conserve Plus	[50]	Wright Medical Technology	USA	100
Elite Ogee	[73]	DePuy	UK	218
Exeter	[87]	n.s.	Norway	3934
Exeter All-Polyethylene	[51]	Stryker	UK	263
Fitmore	[52]	Sulzer	UK	119
Harris Design-2 All-Polyethylene	[89]	Howmedica	Canada	195
Harris-Galante II	[45]	Zimmer	Finland	277
Hofer-Imhof	[90]	Smith and Nephew	Austria	100
Hofer-Imhof	[55]	n.s.	Austria	678
JRI Threaded Cup	[56]	JRI	UK	134
Link	[87]	n.s.	Norway	413
Lord Threaded	[77]	Benoist Girard	Norway	116
Mallory-Head Cementless	[45]	Biomet	Finland	110
MOSC[1]: All Polyethylene	[49]	Biomet	USA	100
MOSC[1]: Metal Backed	[49]	Biomet	USA	134
Morscher Press Fit	[59]	Zimmer	Switzerland	124
Morscher Press Fit	[60]	Sulzer	New Zealand	125
Mueller	[49]	Depuy	USA	141
PCA Pegged	[45]	Howmedica	Finland	122
Romanus	[45]	Biomet	Finland	114
Spectron	[87]	n.s.	Norway	2019
T-28	[64]	Zimmer	USA	132
T-28	[49]	Zimmer	USA	559
Titan	[87]	n.s.	Norway	3205
Trabecular Metal Monoblock ACS[2]	[91]	Zimmer	Greece	156
Universal	[45]	Biomet	Finland	898
Universal	[65]	Biomet	USA	123
Weber Hemispheric	[92]	Hoechst	Netherlands	315
ZA[3] Screw Cup	[93]	Zimmer	Netherlands	135
ZA[3] Screw Cup	[94]	Sulzer	France	200

Table 2.3: **Study Characteristics of all included Studies, describing the Survival Probability of Acetabular Cups for Revision Aseptic Loosening.** n.s.: not specified. 1: Miami Orthopaedic Surgical Clinic. 2: Acetabular Component System. 3: Zweymuller-Alloclassic.

Stem	Ref.	Manufacturer	Country of Study Origin	n at baseline
ABG 1	[45]	Stryker	Finland	390
ABG 1	[46]	Howmedica	UK	100
Anatomic Mesh	[45]	Zimmer	Finland	135
Bimetric	[69]	Biomet	Sweden	104
Bimetric	[45]	Biomet	Finland	1982
Bimetric	[70]	Biomet	USA	105
Charnley	[95]	Thackray	Japan	405
Charnley	[87]	n.s.	Norway	14842
Charnley	[71]	Johnson & Johnson	USA	160
Charnley Elite-Plus	[96]	Depuy	Sweden	114
Charnley Elite-Plus	[97]	Johnson & Johnson	Finland	885
Charnley Flat-back	[97]	Johnson & Johnson	Finland	925
CLS Spotorno	[45]	Sulzer-medica	Finland	108
Corail	[72]	DePuy, France	France	120
Exeter	[87]	n.s.	Norway	3934
Exeter Matt	[97]	Stryker	Finland	876
Exeter Universal	[97]	Stryker	Finland	10620
Exeter Universal	[73]	Howmedica	UK	230
Exeter Universal	[98]	Howmedica	UK	142
Freeman Cemented	[99]	Finsbury Instruments	UK	92
Freeman Cemented	[99]	Finsbury Instruments	UK	97
Freeman Cemented	[100]	Finsbury Instruments	Australia	202
Freeman Cementless	[101]	Finsbury Instruments	UK	100
Freeman Cementless	[74]	Finsbury Instruments	UK	100
Furlong	[56]	JRI	UK	134
Harris Design 2	[89]	Howmedica	Canada	195
Harvard Femoral Stem	[75]	Harvard Health Care	UK	269
Interlok	[97]	Biomet	Finland	581
ITH	[87]	n.s.	Norway	2019
Lord Madreporique	[45]	Benoist Girard	Finland	286
Lord Madreporique	[77]	Benoist Girard	Norway	116
Lubinus IP	[97]	Link	Finland	5790
Lubinus SP II	[97]	Link	Finland	10634
Mallory Head Cementless	[80]	Biomet	USA	2000
Mallory Head Cementless	[58]	Biomet	Canada	307
MS-30	[59]	Zimmer	Switzerland	124
Muller S PCNC[1]	[97]	Zimmer	Finland	2309
Muller S PCNC[1]	[102]	Protek	Switzerland	112
Muller S PCNC[1]	[103]	n.s.	Switzerland	161
Muller S T[2]	[104]	Protek	Germany	203
Muller Style Titanium	[105]	Lima	Slovenia	170
Omnifit Cemented	[71]	Osteonics	USA	305
Osteonics Cemented	[106]	Osteonics	USA	215
Osteonics Cementless	[81]	Stryker	USA	262
PCA	[45]	Howmedica	Finland	111
Profile Porous	[45]	Depuy	Finland	115
R-B Interlok	[71]	Biomet	USA	235
SBG	[82]	Plus Orthopaedics	Austria	230
Stanmore Custom Made	[83]	Depuy	Italy	129
Taperloc	[65]	Biomet	USA	123
Taperloc	[84]	Biomet	USA	129
Titan	[87]	n.s.	Norway	3205
Zweymuller Alloclassic	[94]	Sulzer-medica	France	200

Table 2.4: **Study Characteristics of all included Studies, describing the Survival Probability of Femoral Stems for Revision Aseptic Loosening.** n.s.: not specified. 1: Muller Straight Protasul-10 Cobalt-Nickel-Chromium. 2: Muller Straight Protasul-100 Titanium.

Cup	Ref.	Primary research question	Cohorts construction	Adequacy follow-up	Follow-up performed	n at risk at follow-up	Worst case or comp. risk analysis
ABG 1	[45]	Yes	NC	U	U	≥20	No
ABG 1	[46]	Yes	U	FC	P	<20	Yes
ABG II HA	[44]	Yes	NC	U	P	≥20	No
ACS Triloc+	[47]	Yes	NC	U	U	≥20	No
Arthopor	[47]	Yes	NC	U	U	≥20	No
ATLAS II	[48]	Yes	NC	>5% lost	U	U	No
ATLAS III	[48]	Yes	NC	>5% lost	U	U	No
Charnley	[49]	Yes	NC	U	U	≥20	No
Charnley	[44]	Yes	NC	U	P	≥20	No
Charnley Elite	[44]	Yes	NC	U	P	≥20	No
CLS Spotorno	[44]	Yes	NC	U	P	≥20	No
Conserve Plus	[50]	No	NC	5% lost	P	≥20	No
Exeter All-Poly[1]	[51]	Yes	U	≤5% lost	P	≥20	Yes
Exeter Duration	[44]	Yes	NC	U	P	≥20	No
Fitmore	[52]	Yes	NC	≤5% lost	NP	≥20	No
Harris Design 2	[53]	Yes	C	>5% lost	P	≥20	No
Harris-Galante I	[54]	Yes	NC	≤5% lost	U	≥20	No
Harris-Galante II	[45]	Yes	NC	U	U	≥20	No
Hofer-Imhof	[55]	Yes	U	>5% lost	P	≥20	No
JRI Threaded Cup	[56]	Yes	C	>5% lost	P	≥20	Yes
Lubinus	[44]	Yes	NC	U	P	≥20	No
Lubinus Eccentric	[57]	Yes	U	FC	NP	≥20	No
Mallory-Head[2]	[45]	Yes	NC	U	U	≥20	No
Mallory-Head[2]	[58]	Yes	NC	≤5% lost	P	≥20	No
MOSC: All Poly[3]	[49]	Yes	NC	U	U	≥20	No
MOSC: MB[4]	[49]	Yes	NC	U	U	≥20	No
Morscher Press Fit	[59]	No	NC	FC	P	≥20	No
Morscher Press Fit	[60]	Yes	NC	≤5% lost	P	≥20	No
Mueller	[49]	Yes	NC	U	U	≥20	No
Novae	[61]	Yes	NC	≤5% lost	U	≥20	No
PCA Pegged	[45]	Yes	NC	U	U	≥20	No
Plasmacup	[62]	Yes	NC	>5% lost	NP	≥20	No
RM	[63]	Yes	U	≤5% lost	U	U	Yes
Romanus	[45]	Yes	NC	U	U	≥20	No
T28	[49]	Yes	NC	U	U	≥20	No
T28	[64]	Yes	NC	≤5% lost	P	≥20	No
Trilogy HA	[44]	Yes	NC	U	P	≥20	No
Universal	[45]	Yes	NC	U	U	≥20	No
Universal	[65]	Yes	NC	≤5% lost	P	≥20	No
ZA Screw Cup[5]	[66]	Yes	NC	>5% lost	NP	≥20	No
ZA Screw Cup[5]	[67]	No	U	≤5% lost	U	≥20	No

Table 2.5: **Study Quality of all included Studies, describing the Survival Probability of Acetabular Cups for Revision Any Reason.** U: Unknown. C: Consecutively, NC: Non-Consecutively. FC: Fully Completed. P: Predefined, NP: Non-Predefined.
1: Exeter All-Polyethylene; 2: Mallory-Head Cementless; 3: Miami Orthopaedic Surgical Clinic, All Polyethylene; 4: Miami Orthopaedic Surgical Clinic, Metal Backed; 5: Zweymuller-Alloclassic Screw Cup.

Stem	Ref.	Primary research question	Cohorts construction	Adequacy follow-up	Follow-up performed	n at risk at follow-up	Worst case or comp. risk analysis
ABG 1	[45]	Yes	NC	U	NP	≥20	No
ABG 1	[46]	Yes	U	FC	P	<20	Yes
ABG 1	[44]	Yes	NC	U	P	≥20	No
Anatomic Mesh	[45]	Yes	NC	U	NP	≥20	No
Bicontact	[68]	Yes	U	≤5% lost	NP	≥20	No
Bi-Metric	[69]	Yes	U	≤5% lost	NP	≥20	No
Bi-Metric	[45]	Yes	NC	U	NP	≥20	No
Bi-Metric	[70]	Yes	U	≤5% lost	P	≥20	Yes
Charnley	[71]	Yes	C	>5% lost	NP	≥20	Yes
Charnley	[44]	Yes	NC	U	P	≥20	No
CLS Spotorno	[45]	Yes	NC	U	NP	≥20	No
CLS Spotorno	[44]	Yes	NC	U	P	≥20	No
Conserve Plus	[50]	Yes	NC	≤5% lost	P	≥20	No
Corail	[72]	Yes	U	≤5% lost	NP	≥20	No
Exeter Polished	[44]	Yes	NC	U	P	≥20	No
Exeter Universal	[73]	Yes	U	≤5% lost	P	≥20	Yes
Freeman Cementless	[74]	Yes	U	>5% lost	P	≥20	Yes
Furlong	[56]	Yes	C	≤5% lost	P	≥20	Yes
Harvard Femoral Stem	[75]	Yes	U	≤5% lost	NP	≥20	No
Heritage	[76]	Yes	NC	U	P	≥20	No
Iowa polished	[76]	Yes	NC	U	P	≥20	No
Lord Madreporique	[45]	Yes	NC	U	NP	≥20	No
Lord Madreporique	[77]	Yes	C	≤5% lost	NP	≥20	No
Lubinus IP	[78]	Yes	NC	≤5% lost	P	≥20	No
Lubinus IP	[57]	Yes	NC	FC	NP	≥20	No
Lubinus SP	[78]	Yes	NC	>5% lost	P	≥20	No
Lubinus SP	[57]	Yes	NC	FC	NP	≥20	No
Lubinus SPII	[44]	Yes	NC	U	P	≥20	No
Mallory Head Cemented	[79]	Yes	NC	≤5% lost	NP	≥20	No
Mallory Head Cementless	[80]	Yes	U	>5% lost	NP	≥20	No
Mallory Head Cementless	[58]	Yes	NC	≤5% lost	P	≥20	No
MS-30	[59]	Yes	U	FC	P	≥20	No
Omnifit Cemented	[71]	Yes	NC	>5% lost	NP	≥20	Yes
Osteonics Cementless	[81]	Yes	U	U	NP	≥20	No
PCA	[45]	Yes	NC	U	NP	≥20	No
Profile Porous	[45]	Yes	NC	U	NP	≥20	No
R-B Interlok	[71]	Yes	NC	>5% lost	NP	≥20	Yes
SBG	[82]	Yes	U	>5% lost	NP	≥20	No
Stanmore Custom Made	[83]	Yes	NC	FC	P	≥20	No
Taperloc	[84]	Yes	U	≤5% lost	P	≥20	Yes
Triumph	[76]	Yes	NC	U	P	≥20	No
Zweymuller SL	[66]	Yes	NC	>5% lost	NP	≥20	No

Table 2.6: **Study Quality of all included Studies, describing the Survival Probability of Femoral Stems for Revision Any Reason.** U: Unknown. C: Consecutively; NC: Non-Consecutively. FC: Fully Completed. P: Predefined; NP: Non-Predefined.

Cup	Ref.	Primary research question	Cohorts construction	Adequacy follow-up	Follow-up performed	n at risk at follow-up	Worst case or comp. risk analysis
ABG 1	[45]	Yes	NC	U	U	≥20	No
ACS Triloc+	[47]	Yes	NC	U	U	≥20	No
Arthopor	[47]	Yes	NC	U	U	≥20	No
Brunswik	[85]	Yes	NC	FC	NP	≥20	No
Charnley	[86]	No	C	≤5% lost	NP	≥20	No
Charnley	[85]	Yes	NC	FC	NP	≥20	No
Charnley	[87]	Yes	NC	≤5% lost	NP	≥20	No
Charnley	[49]	Yes	NC	U	U	≥20	No
Charnley All-Poly	[88]	No	NC	>5% lost	U	≥20	No
Conserve Plus	[50]	No	NC	≤5% lost	P	≥20	No
Elite Ogee	[73]	No	NC	≤5% lost	U	U	Yes
Exeter	[87]	Yes	NC	≤5% lost	NP	≥20	No
Exeter All-Poly	[51]	Yes	U	≤5% lost	P	≥20	Yes
Fitmore	[52]	Yes	NC	≤5% lost	NP	≥20	No
Harris Design-2 All-Polyethylene	[89]	Yes	U	≤5% lost	P	≥20	No
Harris-Galante II	[45]	Yes	NC	U	U	≥20	No
Hofer-Imhof	[90]	Yes	NC	FC	P	≥20	No
Hofer-Imhof	[55]	Yes	U	>5% lost	P	≥20	No
JRI Threaded Cup	[56]	Yes	C	>5% lost	P	≥20	Yes
Link	[87]	Yes	NC	≤5% lost	NP	≥20	No
Lord Threaded	[77]	Yes	C	U	P	≥20	Yes
Mallory-Head Cementless	[45]	Yes	NC	U	U	≥20	No
MOSC[1]: All Poly	[49]	Yes	NC	U	U	≥20	No
MOSC[1]: Metal Backed	[49]	Yes	NC	U	U	≥20	No
Morscher Press Fit	[59]	No	NC	FC	P	≥20	No
Morscher Press Fit	[60]	Yes	NC	≤5% lost	P	≥20	No
Mueller	[49]	Yes	NC	U	U	≥20	No
PCA Pegged	[45]	Yes	NC	U	U	≥20	No
Romanus	[45]	Yes	NC	U	U	≥20	No
Spectron	[87]	Yes	NC	≤5% lost	NP	≥20	No
T-28	[64]	Yes	NC	≤5% lost	P	≥20	No
T-28	[49]	Yes	NC	U	U	≥20	No
Titan	[87]	Yes	NC	≤5% lost	NP	≥20	No
Trabecular Metal Monoblock ACS[2]	[91]	Yes	NC	FC	NP	≥20	No
Universal	[45]	Yes	NC	U	U	≥20	No
Universal	[65]	Yes	NC	≤5% lost	P	≥20	No
Weber Hemispheric	[92]	Yes	U	>5% lost	U	≥20	Yes
ZA[3] Screw Cup	[93]	Yes	U	≤5% lost	P	≥20	Yes
ZA[3] Screw Cup	[94]	Yes	NC	U	P	≥20	Yes

Table 2.7: **Study Quality of all included Studies, describing the Survival Probability of Acetabular Cups for Revision Aseptic Loosening.** U: Unknown. C: Consecutively; NC: Non-Consecutively. FC: Fully Completed. P: Predefined; NP: Non-Predefined.
1: Miami Orthopaedic Surgical Clinic. 2: Acetabular Component System. 3: Zweymuller-Alloclassic.

Stem	Ref.	Primary research question	Cohorts construction	Adequacy follow-up	Follow-up performed	n at risk at follow-up	Worst case or comp. risk analysis
ABG 1	[45]	Yes	NC	U	NP	≥20	No
ABG 1	[46]	Yes	U	FC	P	<20	Yes
Anatomic Mesh	[45]	Yes	NC	U	NP	≥20	No
Bimetric	[69]	Yes	U	≤5% lost	NP	≥20	No
Bimetric	[45]	Yes	NC	U	NP	≥20	No
Bimetric	[70]	Yes	U	≤5% lost	P	≥20	Yes
Charnley	[95]	No	U	U	NP	≥20	No
Charnley	[87]	Yes	NC	U	P	≥20	No
Charnley	[71]	Yes	C	>5% lost	NP	≥20	No
Charnley Elite-Plus	[96]	Yes	NC	U	P	≥20	No
Charnley Elite-Plus	[97]	Yes	NC	U	NP	≥20	No
Charnley Flat-back	[97]	Yes	NC	U	NP	≥20	No
CLS Spotorno	[45]	Yes	NC	U	NP	≥20	No
Corail	[72]	Yes	U	≤5% lost	NP	≥20	No
Exeter	[87]	Yes	NC	U	P	≥20	No
Exeter Matt	[97]	Yes	NC	U	NP	≥20	No
Exeter Universal	[97]	Yes	NC	U	NP	≥20	No
Exeter Universal	[73]	Yes	U	≤5% lost	P	≥20	Yes
Exeter Universal	[98]	Yes	U	FC	NP	≥20	No
Freeman Cemented	[99]	No	U	≤5% lost	P	≥20	No
Freeman Cemented	[99]	No	U	≤5% lost	P	≥20	No
Freeman Cemented	[100]	Yes	U	≤5% lost	NP	<20	No
Freeman Cementless	[101]	Yes	U	≤5% lost	P	≥20	Yes
Freeman Cementless	[74]	Yes	U	>5% lost	P	≥20	Yes
Furlong	[56]	Yes	C	≤5% lost	P	≥20	No
Harris Design 2	[89]	Yes	U	≤5% lost	P	≥20	No
Harvard Femoral Stem	[75]	Yes	U	≤5% lost	NP	≥20	No
Interlok	[97]	Yes	NC	U	NP	≥20	No
ITH	[87]	Yes	NC	U	P	≥20	No
Lord Madreporique	[45]	Yes	NC	U	NP	≥20	No
Lord Madreporique	[77]	Yes	C	≤5% lost	NP	≥20	No
Lubinus IP	[97]	Yes	NC	U	NP	≥20	No
Lubinus SP II	[97]	Yes	NC	U	NP	≥20	No
Mallory Head Cementless	[80]	Yes	U	>5% lost	NP	≥20	No
Mallory Head Cementless	[58]	Yes	NC	≤5% lost	P	≥20	No
MS-30	[59]	Yes	U	FC	P	≥20	No
Muller Straight CNC[1]	[97]	Yes	NC	U	NP	≥20	No
Muller Straight CNC[1]	[102]	Yes	NC	≤5% lost	NP	≥20	No
Muller Straight CNC[1]	[103]	Yes	U	≤5% lost	P	≥20	No
Muller Straight Ti[2]	[104]	Yes	NC	>5% lost	NP	≥20	No
Muller Style Titanium	[105]	Yes	NC	≤5% lost	P	≥20	No
Omnifit Cemented	[71]	Yes	C	>5% lost	NP	≥20	No
Osteonics Cemented	[106]	Yes	C	>5% lost	NP	≥20	No
Osteonics Cementless	[81]	Yes	U	U	NP	≥20	No
PCA	[45]	Yes	NC	U	NP	≥20	No
Profile Porous	[45]	Yes	NC	U	NP	≥20	No
R-B Interlok	[71]	Yes	C	>5% lost	NP	≥20	No
SBG	[82]	Yes	U	>5% lost	NP	≥20	No
Stanmore Custom Made	[83]	Yes	NC	FC	P	≥20	No
Taperloc	[65]	Yes	NC	≤5% lost	P	≥20	No
Taperloc	[84]	Yes	U	≤5% lost	P	≥20	Yes
Titan	[87]	Yes	NC	U	P	≥20	No
Zweymuller Alloclassic	[94]	Yes	NC	U	P	≥20	Yes

Table 2.8: **Study Quality of all included Studies, describing the Survival Probability of Femoral Stems for Revision Aseptic Loosening.** 1: Muller Straight Protasul-10 Cobalt-Nickel-Chromium; 2: Muller Straight Protasul-100 Titanium.

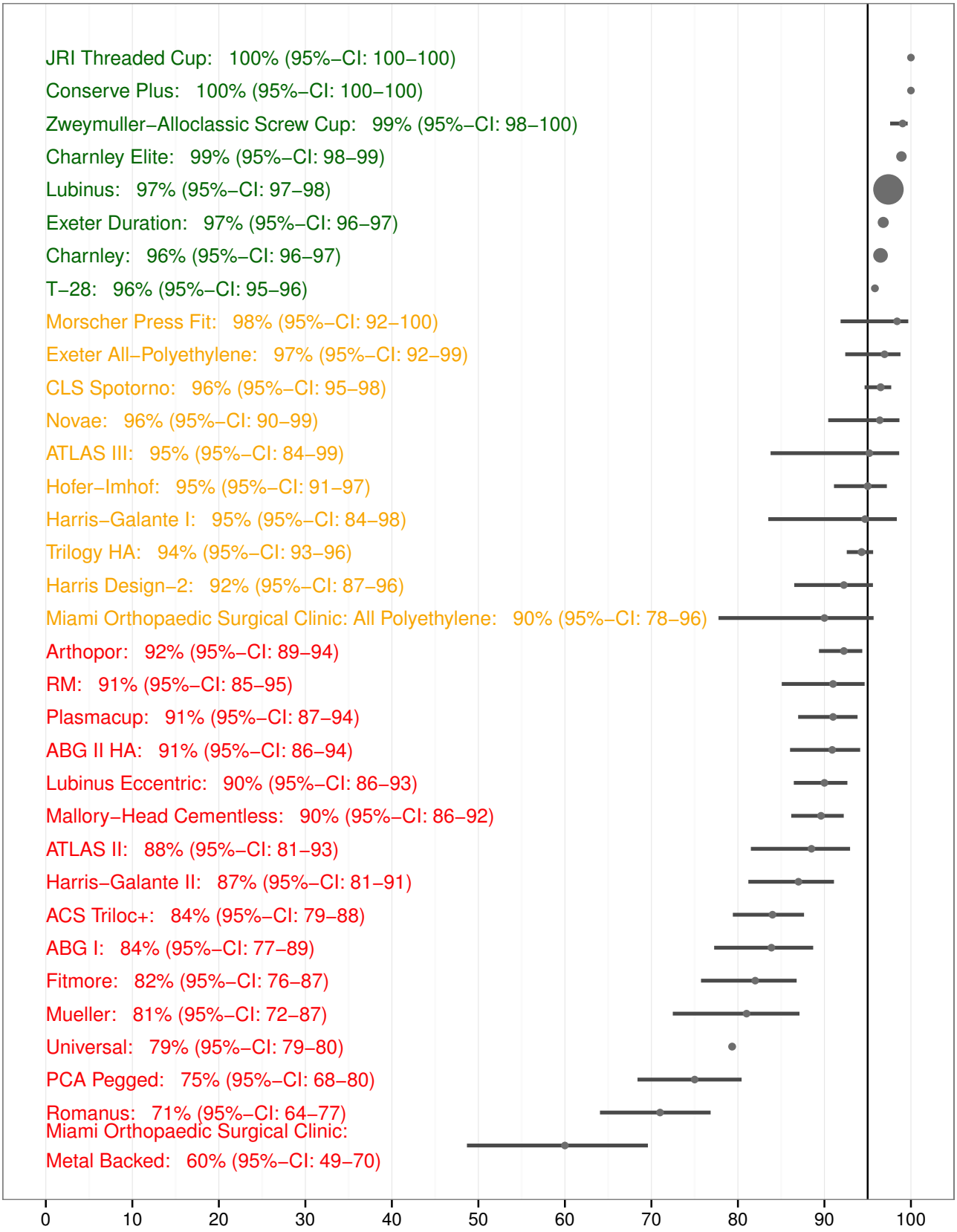


Figure 2.2: **Cumulative Survival and 95% Confidence Intervals of Acetabular Cups at 10 years follow-up, using the endpoint Revision for Any reason.** The vertical line indicates the NICE benchmark; the color of the text indicates whether an implant performs significantly better (green) or worse (red) than the NICE benchmark. The size of the points indicates the sample size on which the estimates are based.

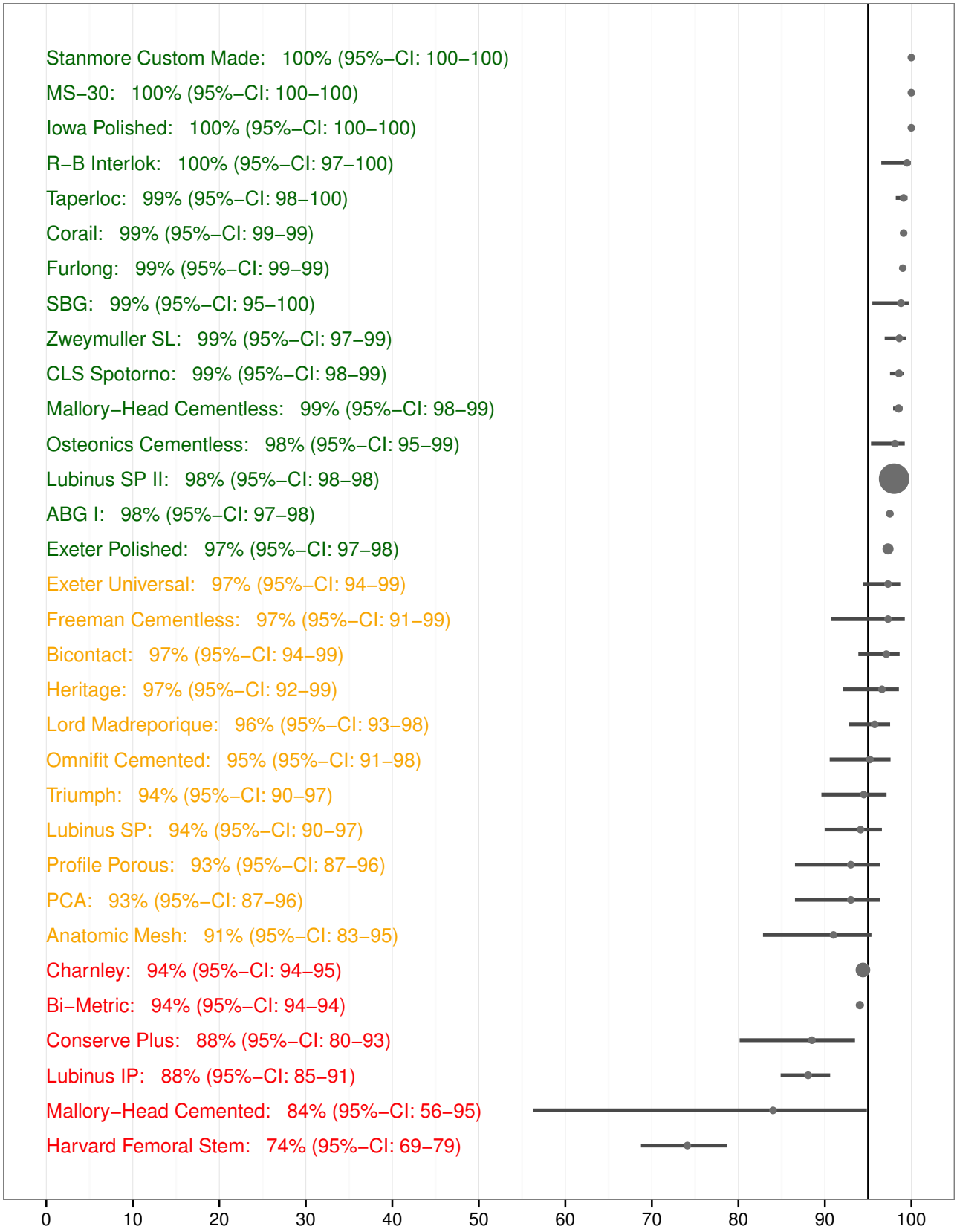


Figure 2.3: **Cumulative Survival and 95% Confidence Intervals of Femoral Stems at 10 years follow-up, using the endpoint Revision for Any reason.** The vertical line indicates the NICE benchmark; the color of the text indicates whether an implant performs significantly better (green) or worse (red) than the NICE benchmark. The size of the points indicates the sample size on which the estimates are based.

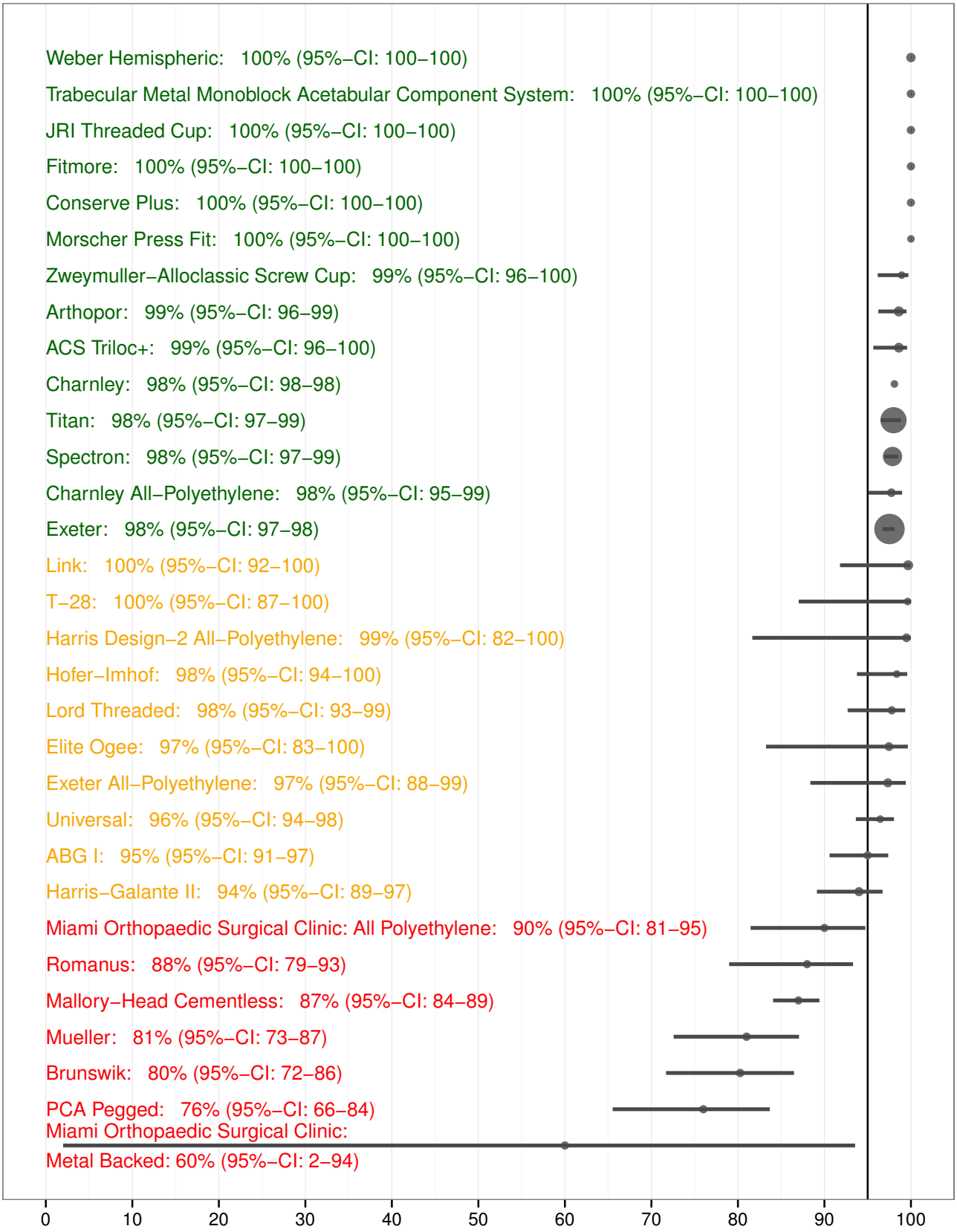


Figure 2.4: **Cumulative Survival and 95% Confidence Intervals of Acetabular Cups at 10 years follow-up, using the endpoint Revision for Aseptic Loosening.** The vertical line indicates the NICE benchmark; the color of the text indicates whether an implant performs significantly better (green) or worse (red) than the NICE benchmark. The size of the points indicates the sample size on which the estimates are based.

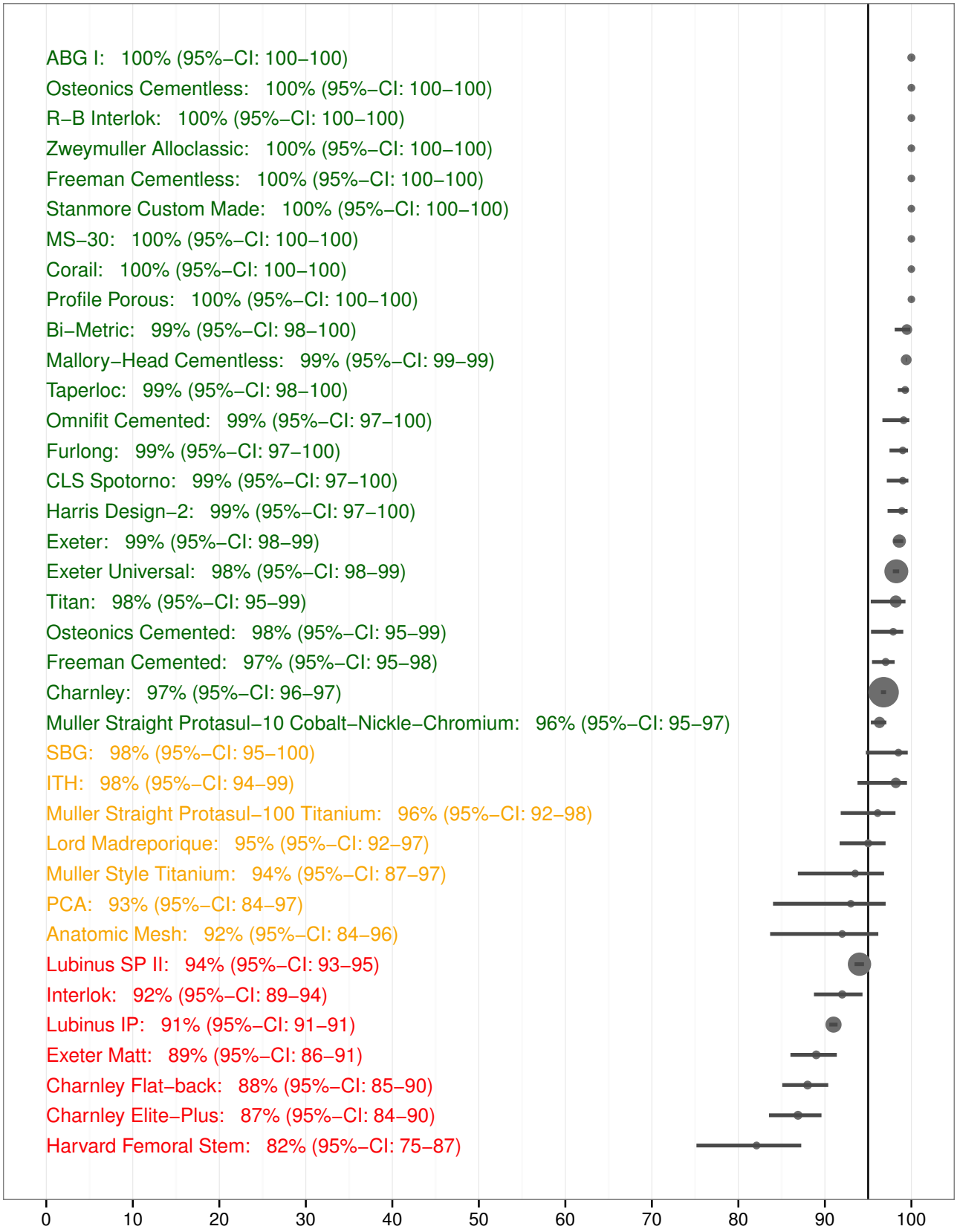


Figure 2.5: **Cumulative Survival and 95% Confidence Intervals of Femoral Stems at 10 years follow-up, using the endpoint Revision for Aseptic Loosening.** The vertical line indicates the NICE benchmark; the color of the text indicates whether an implant performs significantly better (green) or worse (red) than the NICE benchmark. The size of the points indicates the sample size on which the estimates are based.

Part II

Patient and Surgeon Factors

3

Socio-Economic Position Has No Effect on Improvement in Health-Related Quality of Life and Patient Satisfaction in Total Hip and Knee Replacement

JC Keurentjes[1], D Blane[2], M Bartley[3], JJB Keurentjes[4], M Fiocco[5], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Primary Care and Public Health, Imperial College London.

3 Epidemiology and Public Health, University College London.

4 Information Management, Dutch Land Registry Office.

5 Medical Statistics and BioInformatics, Leiden University Medical Center.

PLOS ONE 8 (3), e56785.

Abstract

Introduction Considerable evidence suggests that patients with more advantaged Socio-Economic Positions undergo Total Hip and Knee Replacement (THR/TKR) more often, despite having a lower need. We questioned whether more disadvantaged Socio-Economic Position is associated with an lower improvement in Health-Related Quality of Life (HRQoL) and a lower patient satisfaction after THR/TKR.

Methods Patients who underwent primary THR/TKR in one academic and three community hospitals between 2005 and 2009, were eligible for inclusion. The highest completed levels of schooling were aggregated to index social class. We compared the improvement in HRQoL and postoperative satisfaction with surgery (measured using the Short-Form 36 (SF36) and an 11-point numeric rating scale of satisfaction) between the aggregated groups of highest completed levels of schooling, using linear mixed model analysis, with center as a random effect and potential confounders (i.e. age, gender, Body Mass Index and Charnley's comorbidity classification) as fixed effects.

Results 586 THR patients and 400 TKR patients (40% of all eligible patients) agreed to participate and completed all questionnaires sufficiently. We found no differences in HRQoL improvement in any dimension of the SF36 in THR patients. Patients with a higher completed level of schooling had a larger improvement in role-physical (9.38 points, 95%-CI:0.34–18.4), a larger improvement in general health (3.67 points, 95%-CI:0.56–6.79) and a smaller improvement in mental health (3.60 points, 95%-CI:0.82–6.38) after TKR. Postoperative patient satisfaction did not differ between different highest completed level of schooling groups.

Discussion Completed level of schooling has no effect on the improvement in HRQoL and patient satisfaction in a Dutch THR population and a small effect in a similar TKR population. Undertreatment of patients with more disadvantaged Socio-Economic Position cannot be justified, given the similar improvement in HRQoL and postoperative level of satisfaction with surgery between the social groups examined.

Introduction

Total Hip Replacement (THR) and Total Knee Replacement (TKR) are effective surgical interventions, which alleviate pain and improve Health-Related Quality of Life (HRQoL) in patients with hip or knee joint degeneration.[1] Although on average patients improve markedly after THR or TKR, not all patients benefit from these surgeries. Persistent pain is reported in 9% of THR patients and 20% of TKR patients at long term follow-up.[11] Additionally, up to 30% of patients are dissatisfied after surgery, with higher reported dissatisfaction rates for TKR patients.[12–18] Therapeutic options are limited in patients with persistent pain or dissatisfaction after joint replacement: the outcome of revision surgery performed without a specific mechanical or physiological indication is highly unpredictable. Furthermore, revision THR or TKR surgery is associated with a higher probability of orthopaedic and medical complications. Given the projected increase of 137% and 601% in the annual number of THR and TKR performed in the United States in 2030, the absolute number of patients with unfavorable outcomes after joint replacement is expected to rise, potentially inducing large societal and medical problems.[123]

Predicting which patient groups are at increased risk of an unfavourable outcome after joint replacement may provide additional insights in the mechanisms involved and offer the possibility of intervention in order to optimise the outcome. At the very least, it allows patients to be well informed of their specific risks and expected gains before surgery.

People attain unequal societal positions according to their occupation, educational achievement, income level and status. The Socio-Economic Position (SEP) encompasses both resource-based measures and prestige-based measures in determining an individuals position in the socioeconomic hierarchy.[124] The patients SEP might be a good predictor of a favorable outcome after joint replacement: a more advantaged SEP is associated with better health,[125] which in turn is associated with better outcomes after joint replacement surgery.[126, 127] As it does not require any

invasive or expensive diagnostics, it would be easy to implement in clinical practice. We therefore questioned whether SEP was associated with the improvement in HRQoL and satisfaction after THR or TKR. We hypothesised that patients with more advantaged SEP would have a larger improvement in HRQoL after THR and TKR and a higher degree of satisfaction with their surgical results.

Methods

The presently reported study is an add-on to a multi-center follow-up study, conducted at the departments of orthopaedic surgery of the Leiden University Medical Center, the Slotervaart hospital in Amsterdam, the Albert Schweitzer hospital in Dordrecht and the Groene Hart hospital in Gouda, the Netherlands, from August 2010 until August 2011 (see Study Time-line in figure 3.1 on the facing page). The study was approved by the Medical Ethics Committee of the Leiden University Medical Center and the Medical Ethical Committees of all other participating centers; all patients gave written informed consent (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189). This study was registered in the Netherlands Trial Register (NTR2190). It concerned the clinical follow-up of a multi-center randomized controlled clinical trial, comparing different blood management modalities in THR and TKR surgery (Netherlands Trial Register: NTR303). In this trial, 2442 primary and revision hip or knee replacements in 2257 patients were included between 2004 and 2009 (see Study time line in figure 3.1 on the next page). All patients who participated in the randomized controlled trial and completed preoperative HRQoL questionnaires, who underwent primary THR of TKR and who were alive at the time of inclusion for the present follow-up study were eligible for inclusion. In contrast to the previous clinical trial, in which joint replacements were the subjects of interest, patients are the subject of interest in the current study. Patients who participated more than once in the previous trial, were only allowed to participate once in the current study; the first joint replacement performed in the previous trial was chosen as the index surgery. Records of the financial administration of all participating centers were checked in order to ascertain that all eligible patients were still alive before being approached. All eligible

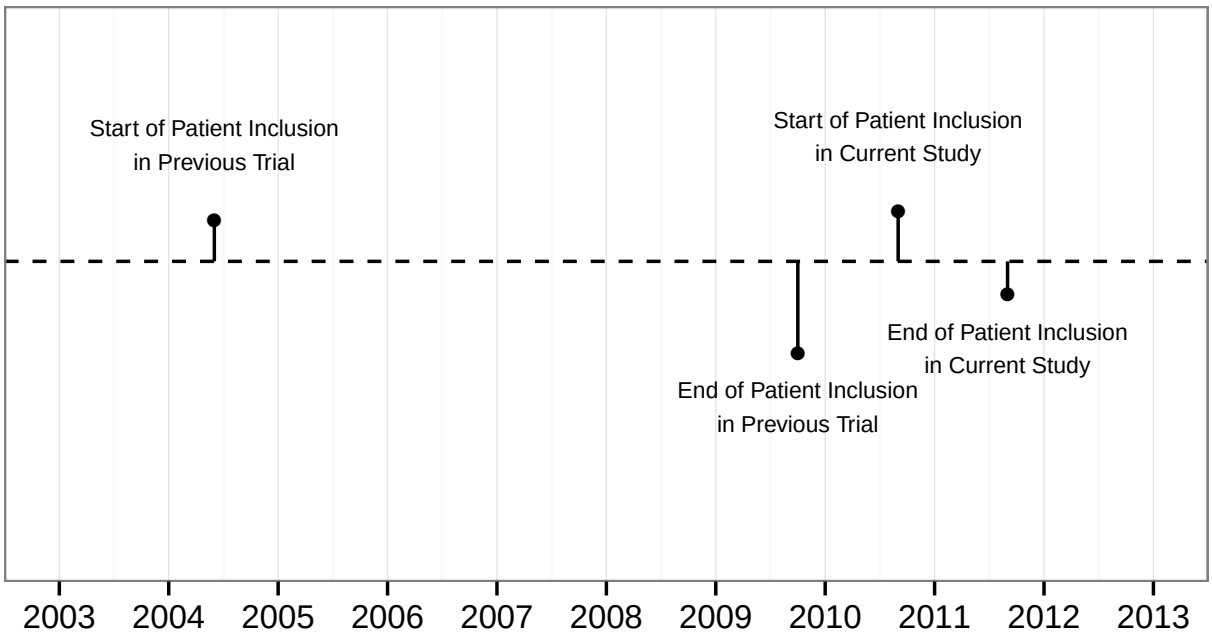


Figure 3.1: **Study time line.**

patients were first sent an invitation letter signed by their treating orthopaedic surgeon, an information brochure and a reply card. Patients who did not respond within 4 weeks after the first invitation were sent another invitation letter. The remaining patients, who did not respond to this second invitation, were contacted by telephone.

Outcome The improvement in different dimensions of HRQoL and satisfaction with the surgical results were the outcome measures of interest. Important concepts in HRQoL are elements of health status that people usually value (e.g. stair climbing) and peoples rating of the value of their subjective experience of living.[128] In other words, both objective functioning and subjective well-being should be considered when measuring HRQoL.[129] We measured HRQoL preoperatively and in the present follow-up study using the Short-Form 36 (SF36),[130] a health status instrument which includes several sub-scales related to functioning as well as perceived well-being.[131, 132] The SF36 is translated and validated in Dutch and allows studying small between-group differences in HRQoL.[133, 134] The 36 items cover eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health), for which a sub-scale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). Additionally, these scales are incorporated into two summary

measures: a Physical Component Summary (PCS) and Mental Component Summary (MCS). The HRQoL outcome measure was the mean improvement (i.e. the mean of each patients postoperative sub-scale score minus their preoperative sub-scale score). At follow-up one question was asked about satisfaction with the result in general, namely: “How satisfied are you with your hip or knee replacement?”. Such as a single item has been shown to provide additional insight into the impact of surgery, besides the measurement of HRQoL.[135] Patient satisfaction with the surgical result was measured using an 11-point Numeric Rating Scale of Satisfaction (NRSS; 0 indicating completely dissatisfied, 10 indicating completely satisfied). The satisfaction outcome measure was the mean NRSS score.

Exposure The follow-up questionnaire contained the following question: “What is your highest completed level of schooling?”. We have aggregated these levels of schooling into an approximation of the social classes, on the assumption that level of schooling indexes the type of qualifications obtained, which in turn indicates the type of occupations available to the subject and hence their own adult social class. Thus: University, Higher vocational education and Preparatory higher vocational & scientific education have been aggregated as indicating the professional and managerial social classes; Middle vocational education and Preparatory middle vocational education have been aggregated as indicating the skilled non-manual and manual social classes; and Lower vocational education, Elementary schooling and No formal education have been aggregated as indicating the semi- & unskilled manual social classes.

Potential confounders Socio-demographic characteristics collected at baseline in the trial included: age at joint replacement and gender. Additionally, the following variables were collected in the questionnaire of the follow-up study: length and weight, in order to calculate the Body Mass Index (BMI) (<25, 25–30, 30–35, >35) and patient reported Charnley classification of co-morbidity (Class A: patients in which the index operated hip or knee are affected only; Class B: patients in which the other hip or knee is affected as well; Class C: patients with a hip or knee replacement and other affected joints and/or a

medical condition which affects the patients' ability to ambulate).[136, 137]

Statistical analysis We performed descriptive analyses of patients baseline characteristics. In order to investigate the possible extent of self-selection bias, we compared the age at THR or TKR and gender of participants to non-participants.

Patients with missing preoperative SF36 questionnaires, missing SF36 questionnaires at follow-up or missing highest level of schooling were excluded from analyses, as we could not exclude a Missing Not At Random (MNAR) mechanism. Missing values of the Charnley Co-morbidity Classification and BMI were deemed Missing At Random and imputed using Multiple Imputations (MI), in order to improve efficiency of the regression analyses and avert biased regression coefficients. We performed MI ($m = 10$) using an Expectation-Maximization algorithm,[138] which is implemented in the Amelia 2 package for R.[139, 140]

We performed regression analyses in each imputed dataset in order to compare the mean improvement in HRQoL and the mean NRSS between patients from different social classes, whilst adjusting for confounders. As minimal clinically important differences (MCIDs) in HRQoL differ between THR patients and TKR patients,[141] we performed all analyses separately for THR and TKR. Possible confounders are age, gender, BMI and poly-articular morbidity in both THR and TKR patients. We used the Charnley classification as a proxy for poly-articular morbidity. As the length of follow-up varies considerably, we first stratified our data in quartiles of follow-up length for each imputed dataset. Within each stratum of follow-up length, we performed a multivariate mixed effect linear regression analysis, with the mean improvement in HRQoL and the mean NRSS as the dependent variable, the completed level of schooling and confounders as independent variables and center as a random effect. Stratum-specific mean differences in HRQoL between the KL grades were pooled using inverse variance weighting in order to produce an overall estimate of the mean difference in HRQoL for each imputed dataset. Finally, the $m = 10$ estimates of the mean differences in HRQoL were combined into one estimate, according to Rubin.[142]

All analyses were performed using R, version 2.14.0.[43]

	U+HVE+PHVSE: n = 100	MVE+PMVE: n = 150	LVE+ES+NFE: n = 156	All Patients: n = 406
Age:	62.5 (11.9)	63.8 (10.6)	66.3 (9.4)	64.4 (10.6)
Males:	51.0%	34.7%	35.9%	39.2%
Follow-up period:	3.13 (1.20)	3.19 (1.10)	3.17 (1.10)	3.16 (1.14)
Charnley Class A:	24.2%	20.0%	22.6%	22.0%
Charnley Class B:	12.6%	17.9%	11.6%	14.2%
Charnley Class C:	63.2%	62.1%	65.8%	63.7%
BMI <25:	50.0%	33.8%	27.0%	35.0%
BMI 25–30:	35.1%	41.2%	48.0%	42.4%
BMI 30–35:	12.8%	20.3%	19.1%	18.0%
BMI >35:	2.10%	4.70%	5.90%	4.60%

Table 3.1: **Patient Characteristics of THR Patients.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (SD), unless stated otherwise.

Results

In the previous trial, 2579 THR and TKR were randomised in 2382 patients; 2442 joint replacements were evaluated. The first joint replacements of the 2382 patients consisted of 2206 primary THR and TKR and 176 revision THR and TKR. Of these 2206 patients who underwent primary joint replacement, 285 patients did not complete all preoperative questionnaires and 63 patients died, leaving 1858 patients with primary joint replacement eligible. 986 patients agreed to participate, of which 668 patients had returned all questionnaires sufficiently completed (response rate: 40%, figure 1.1 (p. 5)). Non-responding THR patients were on average 3.95 years older than participants (95%CI: 2.6 – 5.3 years); Non-responding TKR patients were on average 3.31 years older than participants (95%CI: 2.0 – 4.7 years). The proportion of males was similar in participants and non-responders. An overview of the patient characteristics is provided in table 3.1 and 3.2 on the facing page, an overview of preoperative HRQoL is presented in table 3.3 on the next page for THR patients and 3.4 (p. 46) for TKR patients. Data on age, gender, highest completed level of schooling, pre- and postoperative SF36, satisfaction with surgery and length of follow-up was complete for all THR patients and

	U+HVE+PHVSE: n = 42	MVE+PMVE: n = 98	LVE+ES+NFE: n = 122	All Patients: n = 262
Age:	63.7 (12.7)	67.6 (9.0)	69.2 (9.3)	67.7 (10.0)
Males:	40.5%	35.7%	29.5%	33.6%
Follow-up period:	3.25 (1.20)	3.02 (1.00)	3.28 (1.20)	3.18 (1.13)
Charnley Class A:	14,60%	14.0%	11.7%	13.0%
Charnley Class B:	14.6%	10.8%	10.0%	11.0%
Charnley Class C:	70.6%	75.3%	78.3%	76.0%
BMI $<$25:	23.7%	13.7%	15.5%	16.1%
BMI 25–30:	47.4%	56.8%	37.9%	46.6%
BMI 30–35:	23.7%	21.1%	26.7%	24.1%
BMI $>$35:	5.30%	8.40%	19.8%	13.3%

Table 3.2: **Patient Characteristics of TKR patients.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (SD), unless stated otherwise.

	U+HVE+PHVSE: n = 100	MVE+PMVE: n = 150	LVE+ES+NFE: n = 156	All Patients: n = 406
Physical Functioning	43.0 (20.2)	39.1 (21.7)	39.8 (22.5)	40.3 (21.6)
Role-Physical	38.4 (40.7)	31.7 (39.6)	28.6 (38.1)	32.2 (39.4)
Bodily Pain	44.3 (19.3)	41.7 (20.6)	38.4 (20.7)	41.1 (20.4)
General Health	70.0 (19.9)	69.1 (19.4)	67.6 (19.3)	68.7 (19.5)
Vitality	67.2 (20.7)	59.6 (20.6)	59.5 (22.8)	61.4 (21.7)
Social Functioning	69.0 (22.8)	66.2 (26.6)	63.8 (30.6)	66.0 (27.4)
Role Emotional	79.7 (36.4)	71.1 (41.7)	67.1 (41.7)	71.7 (40.6)
Mental Health	78.9 (15.8)	74.0 (18.1)	73.4 (19.6)	75.0 (18.3)
PCS	38.0 (11.1)	38.5 (9.10)	38.8 (9.40)	38.5 (9.70)
MCS	54.8 (9.30)	51.7 (10.9)	50.9 (11.1)	52.2 (10.7)

Table 3.3: **Quality of Life before Total Hip Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (SD).

all TKR patients. In 20 THR patients and 8 TKR patients, the Charnley classification was missing; in 12 THR patients and 13 TKR patients, the BMI was missing.

The mean improvement in HRQoL and mean NRSS per completed level of schooling is shown in table 3.5 (p. 47) for THR patients and table 3.6 (p. 48) for TKR patients.

	U+HVE+PHVSE: n = 42	MVE+PMVE: n = 98	LVE+ES+NFE: n = 122	All Patients: n = 262
Physical Functioning	40.4 (19.4)	41.3 (19.1)	38.4 (22.1)	39.8 (20.6)
Role-Physical	41.7 (41.9)	40.4 (42.3)	38.1 (42.9)	39.5 (42.4)
Bodily Pain	45.5 (19.4)	45.4 (19.7)	42.2 (21.6)	43.9 (20.6)
General Health	62.5 (19.0)	65.2 (18.7)	59.0 (21.1)	61.9 (20.0)
Vitality	63.2 (18.0)	63.1 (21.2)	57.7 (22.3)	60.6 (21.4)
Social Functioning	72.6 (22.1)	72.2 (23.0)	67.3 (26.3)	70.0 (24.5)
Role Emotional	82.5 (33.1)	74.5 (39.4)	62.0 (44.8)	70.0 (41.8)
Mental Health	79.6 (10.4)	76.4 (15.7)	68.1 (20.3)	73.1 (18.0)
PCS	36.8 (11.8)	40.2 (9.20)	40.8 (9.70)	39.9 (10.0)
MCS	55.5 (7.40)	53.0 (9.50)	48.8 (11.2)	51.5 (10.4)

Table 3.4: **Quality of Life before Total Knee Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (SD).

Adjusted differences in improvement in HRQoL and mean NRSS after joint replacement per increasing category of completed level of schooling are shown in table 3.7 (p. 49) for THR patients and table 3.8 (p. 50) for TKR patients. For each increasing completed level of schooling, THR patients improved 0.88 points more in physical functioning, 3.09 points less in role-physical, 0.60 points less in bodily pain, 0.66 points less in general health, 1.44 points less in vitality, 0.12 points more in social functioning, 0.34 points less in role-emotional, 1.35 points less in mental health, 0.17 points less in the physical component summary and 0.80 points less in the mental component summary; however, none of these differences reached statistical significance (table 3.7 (p. 49)). For each increasing completed level of schooling, TKR patients improved 3.64 points more in physical functioning, 9.38 points more in role-physical, 3.68 points more in bodily pain, 3.67 points more in general health, 1.78 points less in vitality, 0.62 points more in social functioning, 3.11 points less in role-emotional, 3.60 points less in mental health, 2.74 points more in the physical component summary and 2.08 points less in the mental component summary; however, only role-physical, general health, mental health, the physical component summary and the mental component summary reached statistical significance (table 3.8 (p. 50)). For each increasing completed level of schooling, the

	U+HVE+PHVSE: n = 100	MVE+PMVE: n = 150	LVE+ES+NFE: n = 156
Physical Functioning	27.8 (23.3 – 32.3)	26.6 (22.4 – 30.7)	24.9 (20.5 – 29.2)
Role-Physical	35.7 (26.6 – 44.7)	40.7 (32.9 – 48.6)	42.3 (35.0 – 49.5)
Bodily Pain	38.0 (33.1 – 42.9)	33.4 (29.3 – 37.6)	38.9 (34.9 – 42.9)
General Health	-1.20 (-4.80 – 2.50)	-0.70 (-3.70 – 2.30)	-0.20 (-3.50 – 3.10)
Vitality	3.40 (0.20 – 6.60)	8.50 (5.80 – 11.3)	6.70 (3.10 – 10.2)
Social Functioning	16.0 (11.3 – 20.7)	18.1 (13.5 – 22.7)	20.0 (15.4 – 24.5)
Role Emotional	5.70 (-2.60 – 13.9)	16.0 (8.80 – 23.2)	11.9 (4.70 – 19.0)
Mental Health	2.10 (-0.60 – 4.80)	6.40 (3.90 – 9.00)	5.90 (3.30 – 8.50)
PCS	12.8 (11.1 – 14.6)	10.8 (9.20 – 12.3)	11.4 (9.80 – 13.0)
MCS	-1.60 (-3.40 – 0.30)	1.50 (-0.10 – 3.00)	0.60 (-1.00 – 2.20)
NRS Satisfaction	8.9 (8.6 – 9.3)	8.7 (8.4 – 9.0)	8.6 (8.3 – 8.9)

Table 3.5: **Improvement in Health-Related Quality of Life and Satisfaction after Hip Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (95%-Confidence Intervals).

NRSS increased 0.1 points for THR patients and 0.0 points for TKR patients. None of these differences reached statistical significance (table 3.7 (p. 49) and 3.8 (p. 50)). Adjusted differences in improvement in HRQoL and mean NRSS after joint replacement between each completed level of schooling category are shown in table 3.9 (p. 51) for THR patients and table 3.10 (p. 52) for TKR patients. The larger improvement in role-physical functioning in patients with a higher level of completed schooling is mainly due to the large difference between patients with Middle Vocational Education or Preparatory Middle Vocational Education and patients with Lower Vocational Education, Elementary Schooling or No Formal Education. The larger improvement in general health is constant across all groups of level of completed schooling. The larger improvement in the Physical Component Summary Scale in patients with a higher level of completed schooling is mainly due to the large difference between patients with Middle Vocational Education or Preparatory Middle Vocational Education and patients with Lower Vocational Education, Elementary Schooling or No Formal Education. The smaller improvement in mental health in patients with a higher level of completed schooling is mainly due to the

	U+HVE+PHVSE: n = 42	MVE+PMVE: n = 98	LVE+ES+NFE: n = 122
Physical Functioning	20.4 (12.4–28.5)	14.0 (8.70–19.3)	10.3 (6.40–14.2)
Role-Physical	31.5 (15.6–47.5)	25.2 (14.4–35.9)	15.6 (7.40–23.7)
Bodily Pain	24.9 (17.6–32.2)	25.5 (19.7–31.2)	21.0 (16.5–25.5)
General Health	4.00 (-0.60–8.60)	-1.60 (-5.20–2.10)	-3.60 (-6.90—0.30)
Vitality	1.30 (-3.60–6.30)	-1.00 (-4.70–2.80)	2.60 (-0.90–6.10)
Social Functioning	11.6 (2.80–20.4)	7.80 (1.80–13.8)	8.80 (4.40–13.2)
Role Emotional	9.50 (-1.50–20.5)	3.40 (-6.30–13.1)	10.9 (2.60–19.2)
Mental Health	3.30 (0.20–6.40)	1.30 (-2.30–4.90)	6.20 (3.40–9.10)
PCS	7.70 (4.20–11.3)	7.00 (4.90–9.00)	4.20 (2.70–5.70)
MCS	-0.40 (-3.00–2.20)	-1.70 (-3.7–0.30)	1.70 (0.00–3.40)
NRS Satisfaction	8.3 (7.6–9.1)	8.1 (7.6–8.6)	7.9 (7.4–8.4)

Table 3.6: **Improvement in Health-Related Quality of Life and Satisfaction after Knee Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. Values are means (95%-Confidence Intervals).

large difference between patients with Middle Vocational Education or Preparatory Middle Vocational Education and patients with Lower Vocational Education, Elementary Schooling or No Formal Education. Finally, the smaller improvement in the Mental Component Summary Scale in patients with a higher level of completed schooling is mainly due to the large difference between patients with Middle Vocational Education or Preparatory Middle Vocational Education and patients with Lower Vocational Education, Elementary Schooling or No Formal Education.

Discussion

Regardless of their completed level of schooling, patients improve in HRQoL and have a high satisfaction after THR. After TKR, we found that patients with higher completed levels of schooling had a larger improvement in role-physical functioning, general health and the Physical Component Summary scale and a smaller improvement in mental health and the Mental Component Summary scale, although the found differences in the SF36 subscales were smaller than recently published within-group MCIDs at two-

	Adjusted difference per increasing Completed Levels of Schooling (95% CI)	p-value
Physical Functioning	-0.88 (-4.14–2.38)	0.59
Role-Physical	3.09 (-2.89–9.07)	0.31
Bodily Pain	0.60 (-2.70–3.89)	0.72
General Health	0.66 (-1.81–3.13)	0.60
Vitality	1.44 (-1.04–3.92)	0.25
Social Functioning	-0.12 (-3.59–3.36)	0.94
Role Emotional	0.34 (-5.31–6.00)	0.90
Mental Health	1.35 (-0.61–3.30)	0.18
Physical Component Summary Scale	0.17 (-1.04–1.38)	0.79
Mental Component Summary Scale	0.80 (-0.42–2.03)	0.20
Numeric Rating Scale of Satisfaction	-0.1 (-0.4–0.1)	0.29

Table 3.7: **Adjusted Difference in Improvement in Health-Related Quality of Life and Satisfaction after Hip Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** Negative values indicate a higher mean improvement in HRQoL after THR in patients with increasing Completed Levels of Schooling. The mean differences between education level are adjusted for age, sex, Body Mass Index and Charnley Classification of Comorbidity and stratified for quartiles of follow-up.

years follow-up.[141] All other dimensions of HRQoL and patient satisfaction showed no differences between the completed levels of schooling, thereby failing to refute our hypothesis.

Strengths of our study include the rigorous efforts to minimise confounding and the generalisability of our study population, due to the multi-center setting and the similarity of the demographics of our study population to those of large-scaled national joint registries.[143]

Weaknesses of the study include the low participation rate and the variation in follow-up period after joint replacement. Although participation rates of 100% are feasible in small-scaled studies with hard endpoints,[34, 116] participation rates in epidemiological studies have been steadily declining in the last 30 years.[144] Even sharper declines have been reported in the past few years.[145] Unfortunately, the participation rate of this study follows this general trend, and therefore we cannot exclude the presence of self-selection bias. In order to limit the extent of this bias, we have sent multiple reminders

	Adjusted difference per increasing Completed Levels of Schooling (95% CI)	p-value
Physical Functioning	-3.64 (-8.03–0.74)	0.10
Role-Physical	-9.38 (-18.4—0.34)	0.04
Bodily Pain	-3.68 (-8.39–1.03)	0.13
General Health	-3.67 (-6.79—0.56)	0.02
Vitality	1.78 (-1.51–5.08)	0.29
Social Functioning	-0.62 (-5.37–4.14)	0.80
Role Emotional	3.11 (-5.07–11.3)	0.46
Mental Health	3.60 (0.82–6.38)	0.01
Physical Component Summary Scale	-2.74 (-4.41—1.07)	0.001
Mental Component Summary Scale	2.08 (0.37–3.79)	0.02
Numeric Rating Scale of Satisfaction	0.0 (-0.5–0.4)	0.83

Table 3.8: **Adjusted Difference in Improvement in Health-Related Quality of Life and Satisfaction after Knee Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** Negative values indicate a higher mean improvement in HRQoL after TKR in patients with increasing Completed Levels of Schooling. The mean differences between education level are adjusted for age, sex, Body Mass Index and Charnley Classification of Comorbidity and stratified for quartiles of follow-up.

and have called all patients who did not answer our reminders and who did not return the questionnaire. As incentives, we have included an appealing information brochure in which the primary goals of the follow-up study were explained and a study pen as a small gift. Additionally, patients were urged to participate by their treating physician. However, the participation rate alone does not determine the extent of bias present in any particular study.[145] The difference between participants and nonparticipants is far more important.[146] As the found differences in demographics were small, it is unlikely that the study results will be severely biased.

The follow-up period after joint replacement varies between 1.5 and 6 years in this study (figure 3.1 (p. 41)). Theoretically, this broad range could influence our findings. In order to exclude this variable, all patients should have been followed for the exact same amount of time. In our data, we found no clear evidence of a relationship between the improvement in HRQoL after joint replacement and the follow-up period (See Appendix S1 and S2 for scatter plots of the improvement in HRQoL as a function of the follow-up period length, stratified per completed levels of schooling and Appendix

	U+HVE+PHVSE vs MVE+PMVE* (95%CI)	MVE+PMVE vs LVE+ES+NFE** (95%CI)	U+HVE+PHVSE vs LVE+ES+NFE*** (95%CI)
Physical Functioning	-1.96 (-8.56–4.64)	0.91 (-5.16–6.97)	-1.94 (-8.53–4.66)
Role-Physical	6.21 (-5.85–18.3)	2.66 (-8.04–13.4)	6.21 (-5.91–18.3)
Bodily Pain	-6.27 (-12.9–0.33)	7.28 (1.69–12.7)	-0.33 (-6.94–6.29)
General Health	-0.70 (-5.72–4.31)	1.57 (-2.95–6.08)	1.16 (-3.85–6.17)
Vitality	5.11 (0.09–10.1)	-1.08 (-5.76–3.61)	3.18 (-1.83–8.19)
Social Functioning	-0.50 (-7.57–6.57)	0.85 (-5.72–7.42)	-0.34 (-7.37–6.70)
Role Emotional	5.63 (-5.81–17.1)	-2.95 (-13.3–7.42)	1.89 (-9.59–13.4)
Mental Health	3.43 (-0.59–7.44)	-0.19 (-3.86–3.48)	2.98 (-0.99–6.95)
PCS	-1.64 (-4.07–0.79)	1.85 (-0.29–3.99)	-0.05 (-2.49–2.38)
MCS	2.69 (0.20–5.19)	-0.70 (-2.95–1.55)	1.93 (-0.55–4.41)
NRS Satisfaction	-0.3 (-0.8–0.2)	0.1 (-0.3–0.6)	-0.3 (-0.9–0.2)

Table 3.9: **Adjusted Difference in Improvement in Health-Related Quality of Life and Satisfaction after Hip Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education.
* Negative values indicate a higher mean improvement in HRQoL after THR in U+HVE+PHVSE patients, compared to MVE+PMVE patients. ** Negative values indicate a higher mean improvement in HRQoL after THR in MVE+PMVE patients, compared to LVE+ES+NFE patients. *** Negative values indicate a higher mean improvement in HRQoL after THR in U+HVE+PHVSE patients, compared to LVE+ES+NFE patients.
The mean differences between education level are adjusted for age, sex, Body Mass Index and Charnley Classification of Comorbidity and stratified for quartiles of follow-up.

S3 and S4 for scatter plots of the NRSS after surgery as a function of the follow-up period length, stratified per completed levels of schooling. All appendices are freely available at: *http://dx.plos.org/10.1371/journal.pone.0056785*). In order to account for this range, we stratified our analysis per quartile of follow-up period. Stratifying for an additional variable inevitably leads to a loss of power, thereby increasing the probability of a type 2-error. In our analysis, this loss of power was negligible, as unstratified analyses showed similar results, supporting our conclusions (data not shown). Although a residual effect of follow-up length within each stratum cannot be excluded, we do not think this is very plausible, as recent evidence suggests that the improvement in HRQoL after completion of the initial rehabilitation-period is sustained up to 7 years after joint

	U+HVE+PHVSE vs MVE+PMVE* (95%CI)	MVE+PMVE vs LVE+ES+NFE** (95%CI)	U+HVE+PHVSE vs LVE+ES+NFE*** (95%CI)
Physical Functioning	-5.80 (-15.1–3.53)	-2.24 (-9.14–4.65)	-7.99 (-17.2–1.26)
Role-Physical	-3.37 (-22.5–15.7)	-12.2 (-26.5–2.18)	-16.5 (-35.4–2.53)
Bodily Pain	1.00 (-9.02–11.0)	-6.46 (-14.1–1.20)	-6.13 (-16.1–3.81)
General Health	-4.84 (-11.4–1.72)	-2.84 (-7.88–2.21)	-7.64 (-14.2—1.10)
Vitality	-0.60 (-7.57–6.36)	3.57 (-1.70–8.83)	2.98 (-3.93–9.89)
Social Functioning	-3.22 (-13.4–6.94)	0.87 (-6.47–8.21)	-2.09 (-12.2–7.94)
Role Emotional	-1.26 (-18.8–16.2)	5.73 (-7.79–19.3)	4.49 (-12.8–21.8)
Mental Health	-0.80 (-6.69–5.08)	5.84 (1.11–10.6)	6.32 (0.53–12.1)
PCS	-1.36 (-4.90–2.18)	-3.42 (-6.04—0.80)	-5.04 (-8.56—1.52)
MCS	0.03 (-3.63–3.70)	3.04 (0.19–5.88)	3.58 (-0.03–7.19)
NRS Satisfaction	0.2 (-0.8–1.1)	-0.1 (-0.8–0.6)	-0.1 (-1.0–0.8)

Table 3.10: **Adjusted Difference in Improvement in Health-Related Quality of Life and Satisfaction after Knee Replacement: A Comparison Between Patients with different Completed Levels of Schooling.** U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education; MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education; LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education. PCS: Physical Component Summary Scale; MCS: Mental Component Summary Scale; NRS: Numeric Rating Scale.
* Negative values indicate a higher mean improvement in HRQoL after TKR in U+HVE+PHVSE patients, compared to MVE+PMVE patients. ** Negative values indicate a higher mean improvement in HRQoL after TKR in MVE+PMVE patients, compared to LVE+ES+NFE patients. *** Negative values indicate a higher mean improvement in HRQoL after TKR in U+HVE+PHVSE patients, compared to LVE+ES+NFE patients.
The mean differences between education level are adjusted for age, sex, Body Mass Index and Charnley Classification of Comorbidity and stratified for quartiles of follow-up.

replacement surgery.[147, 148] The minimum follow-up period is well beyond the length of the expected rehabilitation-period, suggested by a recently published systematic review.[149]

Two other studies have investigated the relation between SEP and patient-reported outcomes after THR or TKR.[150, 151] Allen Butler et. al. have studied this relation in a randomised controlled trial, which compared two THR designs.[150] In this study, the effect of SEP was studied on a multitude of outcome measures, including the WOMAC, Short Form-12 (SF12) and degree of patient satisfaction. An association was found between lower levels of education and a degree of satisfaction which was “less than

very satisfied". Unfortunately, the authors have only reported their significant findings; differences in WOMAC or SF12 between social classes are not reported. A dditionally, only p-values are reported instead of mean differences or relative risks, precluding any judgment on the clinical relevance of their findings. Finally, it is unclear for which factors any associations were adjusted, as the authors applied forward stepwise logistic regression modeling, without mentioning which variables were included in the final model. Davis et. al. have measured WOMAC scores before surgery and at 3, 12 and 24 months after TKR.[151] Whilst comparing WOMAC scores at each time point between patients of different income categories, patients with more disadvantaged SEP had worse preoperative WOMAC scores and similar postoperative WOMAC scores as patients with less disadvantaged SEP. These findings imply a larger improvement in disease-specific quality of life in patients with more disadvantaged SEP than in patients with less disadvantaged SEP. However, not all patients were measured at each time point. A cross-sectional comparison at each time point precludes judgment on the actual within-patient improvement in disease-specific quality of life.
Due to methodological shortcomings of both other studies which investigated the relation between SEP and patient-reported outcomes after joint replacement, no meaningful comparison of results can be made.

Our findings have large implications for policymakers, as a more advantaged SEP is associated with greater use of health services in general.[125] A recent systematic review and numerous studies indicate that this also holds for THR[152–161] and TKR[152, 156–158, 160–162] in post-industrialised countries. Additionally, the need for joint replacement appears to be higher in patients with more disadvantaged SEP,[158, 161, 162] thereby increasing the inequity in access to joint replacement. Undertreatment of patients with more disadvantaged SEP cannot be justified, given the similar improvement in HRQoL and postoperative level of satisfaction with surgery between the examined groups of completed level of schooling.

A number of factors might explain the found differences in improvement in HRQoL after between THR and TKR patients per completed level of schooling groups. Biomechanical factors might play a role. The hip joint is a relatively simple ball and socket joint, which is adequately mimicked by a THR. The adequate mimicry of the biomechanics is reflected in a highly consistent improvement in HRQoL, regardless of completed level of schooling. The biomechanical aspects of the knee joint are more difficult to imitate, as the knee is a pivotal hinge joint with 6 degrees of freedom. These degrees of freedom are generally not restored after TKR, which is substantiated in kinematic and kinetic studies.[163] However, more complex biomechanics might explain a less consistent improvement in HRQoL in TKR patients, but does not explain differences between patient groups with different completed levels of schooling.

Differences between THR and TKR patients might be part of the explanation. Better general health, physical, emotional and social function, motivation and self-efficacy and lower levels of pain before surgery and during the rehabilitation period are associated with improved short- and medium-term outcomes.[11] In our study population, differences in the preoperative health status between completed level of schooling groups are more pronounced in TKR patients than in THR patients (table 3.3 (p. 45) and 3.4 (p. 46)). Finally, differences in rehabilitational options could play an important role. TKR patients require more rehabilitation than THR patients in order to achieve optimal results.[164] TKR patients with higher completed Level of Schooling might have better access to physical therapy or other rehabilitational facilities, and therefore gain more in role-physical functioning and general health than less advantaged patients. This effect might be exacerbated by the higher prevalence of obesity and co-morbidity in TKR patients compared to THR patients. Unfortunately, we do not have any information on the rehabilitational regime of our THR and TKR patients, leaving this hypothesis to be addressed in future research.

4

Patients with Severe Radiographic Osteoarthritis Have a Better Prognosis in Physical Functioning after Hip and Knee Replacement

JC Keurentjes[1], M Fiocco[2], C So-Osman[3], R Onstenk[4], AW Koopman-Van Gemert[5], RG Pöll[6], HM Kroon[7], TPM Vliet Vlieland[1], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Sanquin Blood Supply, South West Region.

4 Orthopaedic Surgery, Groene Hart Hospital, Gouda.

5 Anaesthesiology, Albert Schweitzer Hospital, Dordrecht.

6 Orthopaedic Surgery, Slotervaart Hospital, Amsterdam.

7 Radiology, Leiden University Medical Center.

PLOS ONE 8 (4), e59500.

Abstract

Introduction Although Total Hip and Knee Replacements (THR/TKR) improve Health-Related Quality of Life (HRQoL) at the group level, up to 30% of patients are dissatisfied after surgery due to unfulfilled expectations. We aimed to assess whether the pre-operative radiographic severity of osteoarthritis (OA) is related to the improvement in HRQoL after THR or TKR, both at the population and individual level.

Methods In this multi-center observational cohort study, HRQoL of OA patients requiring THR or TKR was measured 2 weeks before surgery and at 2–5 years follow-up, using the Short-Form 36 (SF36). Additionally, we measured patient satisfaction on a 11-point Numeric Rating Scale (NRSS). The radiographic severity of OA was classified according to Kellgren and Lawrence (KL) by an independent experienced musculoskeletal radiologist, blinded for the outcome. We compared the mean improvement and probability of a relevant improvement (defined as a patients change score $\geq$ Minimal Clinically Important Difference) between patients with mild OA (KL Grade 0–2) and severe OA (KL Grade 3+4), whilst adjusting for confounders.

Results Severe OA patients improved more and had a higher probability of a relevant improvement in physical functioning after both THR and TKR. For TKR patients with severe OA, larger improvements were found in General Health, Vitality and the Physical Component Summary Scale. The mean NRSS was also higher in severe OA TKR patients.

Discussion Patients with severe OA have a better prognosis after THR and TKR than patients with mild OA. These findings might help to prevent dissatisfaction after THR and TKR by means of patient selection or expectation management.

Introduction

Total Hip Replacement (THR) and Total Knee Replacement (TKR) are effective surgical interventions, which alleviate pain and improve Health-Related Quality of Life (HRQoL) in patients with hip or knee joint degeneration at the population level.[1] Although on average patients improve markedly after THR or TKR, not all patients benefit from

these surgeries. Persistent pain is reported in 9% of THR patients and 20% of TKR patients at long term follow-up.[11] Additionally, up to 30% of patients are dissatisfied after surgery, with higher reported dissatisfaction rates for TKR patients.[12–18] The relatively high dissatisfaction rate is especially worrying, as the therapeutic options are limited in dissatisfied patients after joint replacement. Moreover, given the projected increase in the annual number of THR and TKR performed in the United States, the absolute number of dissatisfied patients is expected to rise.[165]

Unattained expectations of surgery are thought to play an important role in dissatisfaction after joint replacement.[12, 13, 15, 166] In order to successfully manage patient expectations, accurate prediction of the probability of a meaningful improvement for each individual patient is of paramount importance. This probability can be assessed at the individual level using the Minimal Clinically Important Difference (MCID), which is defined as the minimal difference in scores of an outcome measure that is perceived by patients as beneficial or harmful.[167, 168] MCIDs in HRQoL, measured using the Short-Form 36, have been established for THR and TKR.[141, 169, 170]

Reports of the effect of the preoperative radiographic severity of osteoarthritis (OA) on the outcome of THR are conflicting: at the population level, Nilsdotter et al showed no effect at one year follow-up, while Meding et al found less postoperative pain at one year follow-up in patients with more preoperative joint space narrowing.[171, 172] At the individual level, patients with severe preoperative radiographic OA were more likely to improve in physical functioning.[173] We found no studies addressing the effect of the preoperative radiographic severity of osteoarthritis (OA) on the outcome of TKR.

From a clinical perspective, the preoperative radiographic severity of OA would be a helpful predictor of improvement in HRQoL, as it is both inexpensive and performed routinely for templating purposes. Moreover, the assessment of the severity of preoperative OA could be standardised, whereas this would be more difficult with subjective symptoms such as pain.

We questioned whether the radiographic severity of OA affects the improvement in HRQoL after THR and TKR, both at the population and individual level. Additionally, we questioned whether patient satisfaction with the surgical results differed between patients with mild or severe preoperative radiographical OA.

Methods

We conducted a multi-center follow-up study at the departments of orthopaedic surgery of the Leiden University Medical Center, the Slotervaart hospital in Amsterdam, the Albert Schweitzer hospital in Dordrecht and the Groene Hart hospital in Gouda, the Netherlands, from August 2010 until August 2011.[21] The study was approved by the Medical Ethics Committee of the Leiden University Medical Center and the Medical Ethical Committees of all other participating centers; all patients gave written informed consent (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189). This study was registered in the Netherlands Trial Register (NTR2190). It concerned the clinical follow-up of a multi-center randomized controlled clinical trial, comparing different blood management modalities in THR and TKR surgery (Netherlands Trial Register: NTR303). In this trial, 2442 primary and revision hip or knee replacements in 2257 patients were included between 2004 and 2009.

All patients who participated in the randomized controlled trial and completed preoperative HRQoL questionnaires, who underwent primary THR of TKR for primary OA and who were alive at the time of inclusion for the present follow-up study were eligible for inclusion. In this study, patients are the subject of interest. Patients who participated more than once in the previous trial, were only allowed to participate once in the current study; the first joint replacement performed in the previous trial was chosen as the index surgery.

Records of the financial administration of all participating centers were checked in order to ascertain that all eligible patients were still alive before being approached. All eligible patients were first sent an invitation letter signed by their treating orthopaedic surgeon, an information brochure and a reply card. Patients who did not respond within 4 weeks

after the first invitation were sent another invitation letter. The remaining patients, who did not respond to this second invitation, were contacted by telephone.

Assessments The assessments of the follow-up study consisted of patient-reported questionnaires, examination of patient records and preoperative radiographs.

Outcomes: HRQoL was measured preoperatively and in the present follow-up study using the SF36, which is translated and validated in the Dutch language.[130, 133] The 36 items cover eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health), for which a sub-scale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). Additionally, these scales are incorporated into two summary measures: a Physical Component Summary (PCS) and Mental Component Summary (MCS).

At the population level, the HRQoL outcome measure was the mean change score, i.e. the mean of each patients postoperative sub-scale score minus their pre-operative sub-scale score). At the individual level, the change scores were used to categorise patients in responders and non-responders, using previously published MCIDs.[141, 169, 170] Patients with a change score equal to or larger than the MCID of that particular sub-scale were categorised as a responder; patients whose change score was less than the CID of that particular sub-scale were categorised as non-responders.

Patient satisfaction with the surgical result was measured using an 11-point Numeric Rating Scale of Satisfaction (NRSS; 0 indicating completely dissatisfied, 10 indicating completely satisfied). At the population level, the satisfaction outcome measure was the mean NRSS score. The proportion of patients who achieved a satisfactory outcome (defined as a NRSS $>$ 8, according to Brokelman et al[14]) was the satisfaction outcome measure at the individual level.

Exposure: Pre-operative radiographs of the hips (anterior–posterior) and knees (posterior–anterior) were collected from the participating patients' medical records and radiology department. These radiographs were routinely made in each participating center for pre-operative templating purposes. All radiographs were assessed by an experienced musculoskeletal radiologist (HMK), who was blinded for patient

characteristics and HRQoL assessments. The method of scoring OA followed that described by Kellgren and Lawrence (KL) (0 indicating no OA, 1 doubtful OA, 2 minimal OA, 3 moderate OA and 4 indicating severe OA).[174] All radiographs were scored twice: both readings were used to establish intra-reader reliability (Intra-Class Correlation hip radiographs: 0.85 (95%CI: 0.82 – 0.88); Intra-Class Correlation knee radiographs: 0.87 (95%CI: 0.83 – 0.89)). The second reading was used for further statistical analyses.
As KL grade 0 to 2 and grade 3 and 4 are deemed similar from a clinical perspective, we grouped the severity of pre-operative OA in 2 categories: mild radiographic OA (KL grade 0, 1 or 2) and severe radiographic OA (KL grade 3 or 4).

Potential confounders: Socio-demographic characteristics collected at baseline in the trial included: age at joint replacement and gender. Additionally, the following socio-demographic variables were collected in the questionnaire of the follow-up study: length and weight, in order to calculate the Body Mass Index (BMI) (<25, 25–30, 30–35, >35) and patient reported Charnley classification of co-morbidity (Class A: patients in which the index operated hip or knee are affected only; Class B: patients in which the other hip or knee is affected as well; Class C: patients with a hip or knee replacement and other affected joints and/or a medical condition which affects the patients' ability to ambulate).[136, 137]

Statistical Analysis We performed descriptive analyses of patients baseline characteristics. In order to investigate the possible extent of self-selection bias, we compared the age at THR or TKR and gender of participants to non-participants.
Patients with missing pre-operative SF36 questionnaires, missing SF36 questionnaires at follow-up or missing pre-operative radiographs were excluded from analyses, as we could not exclude a Missing Not At Random (MNAR) mechanism. Missing values of the Charnley Co-morbidity Classification and BMI were deemed Missing At Random and imputed using Multiple Imputations (MI), in order to improve efficiency of the regression analyses and avert biased regression coefficients. We performed MI ($m = 10$) using an Expectation-Maximization algorithm,[138] which is implemented in the Amelia 2 package for R.[139, 140]

We performed regression analyses in each imputed dataset in order to compare the mean improvement in HRQoL and the probability of achieving a MCID in HRQoL after THR and TKR, between patients with KL grade 0, 1 or 2 and grade 3 or 4. As MCIDs in HRQoL differ between THR patients and TKR patients, we performed all analyses separately for THR and TKR. Possible confounders are age, gender, BMI and poly-articular OA in both THR and TKR patients. We used the Charnley classification as a proxy for poly-articular OA. As the length of follow-up varied considerably, we first stratified our data in quartiles of follow-up length for each imputed dataset. Within each stratum of follow-up length, we performed a multivariate mixed effect linear regression analysis, with the mean improvement in HRQoL and the mean NRSS as the dependent variable, the KL grade and confounders as independent variables and center as a random effect. Stratum-specific mean differences in HRQoL between the KL grades were pooled using inverse variance weighting in order to produce an overall estimate of the mean difference in HRQoL for each imputed data-set. Finally, the $m = 10$ estimates of the mean differences in HRQoL were combined into one estimate, according to Rubin.[142]

Within each stratum of follow-up length, we also performed a multivariate mixed effect logistic regression analysis, with the probability of attaining a MCID in HRQoL and a satisfactory NRSS as the dependent variable, the KL grade and confounders as independent variables and center as a random effect. Stratum-specific odds ratios of attaining a MCID in HRQoL between the KL grades were pooled using inverse variance weighting in order to produce an overall estimate of the odds ratio of attaining a MCID in HRQoL for each imputed data-set. Finally, the $m = 10$ estimates of the mean differences in HRQoL were combined into one estimate, according to Rubin.[142]

All analyses were performed using R, version 2.14.0.[43]

Results

At 2 to 5 years after joint replacement, 723 patients agreed to participate and returned the questionnaires sufficiently completed (participation rate: 46%, figure 4.1 and 1.1

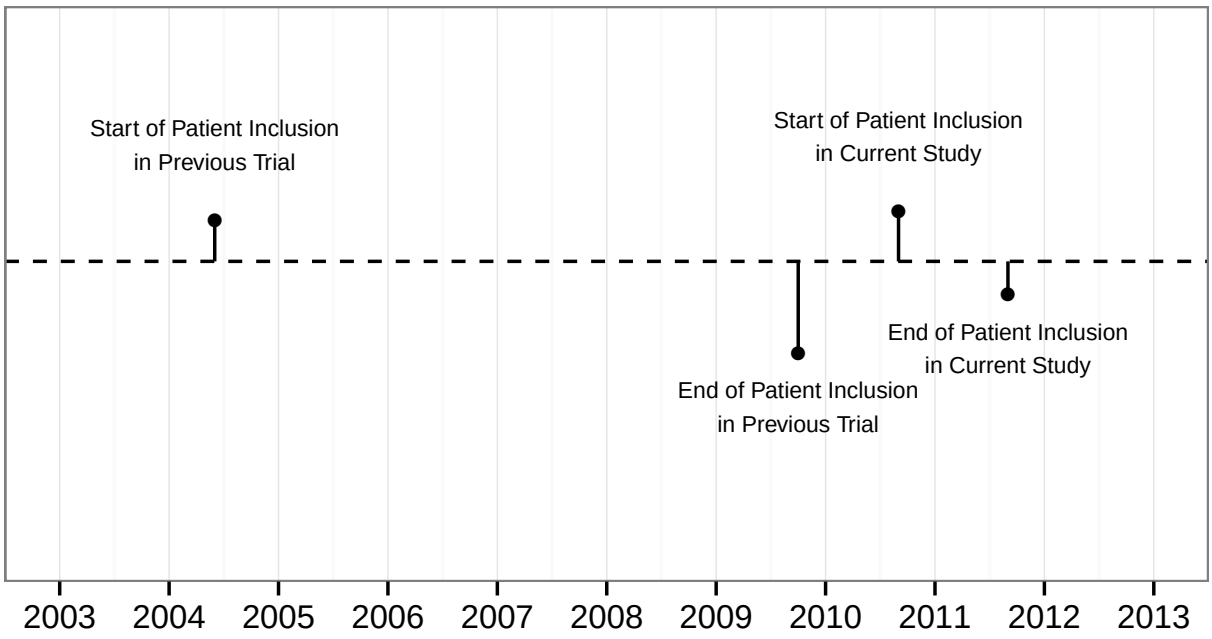


Figure 4.1: **Study time line.**

	Kellgren Grade 0–2	Kellgren Grade 3+4	All Patients
Age at Joint Replacement	65.1 (7.8)	67.4 (8.7)	66.6 (8.5)
Males	23.90%	44.30%	37.50%
Follow-up years (SD)	2.83 (1.0)	2.79 (0.9)	2.8 (0.93)
Charnley Class A:	25.8%	24.0%	24.6%
Charnley Class B:	14.6%	13.7%	14.0%
Charnley Class C:	59.6%	62.3%	61.4%
BMI $<$25:	29.2%	34.1%	32.5%
BMI 25–30:	41.6%	46.4%	44.8%
BMI 30–35:	23.6%	16.2%	18.7%
BMI $>$35:	5.60%	3.40%	4.10%

Table 4.1: **Patient Characteristics of THR Patients.**

(p. 5)). Non-participating THR patients were on average 4.32 years older than participants (95%CI: 2.93 – 5.70 years); Non-participating TKR patients were on average 2.68 years older than participants (95%CI: 1.28 – 4.09 years). The proportion of males was similar in participants and non-responders. An overview of the patient characteristics is provided in table 4.1 and table 4.2 on the facing page. In 13 THR patients and 7 TKR patients, the Charnley classification was missing; in 9 THR patients and 11 TKR patients, the BMI was missing. These missing values were imputed using multiple imputation.

	Kellgren Grade 0–2	Kellgren Grade 3+4	All Patients
Age at Joint Replacement	65.1 (10.3)	69.5 (8.6)	69.1 (8.9)
Males	31.80%	30.90%	31.00%
Follow-up years (SD)	3.1 (1.1)	2.79 (0.9)	2.82 (0.93)
Charnley Class A:	4.50%	19.0%	17.5%
Charnley Class B:	4.50%	11.4%	10.7%
Charnley Class C:	90.9%	69.6%	71.8%
BMI $<$25:	33.3%	14.7%	16.3%
BMI 25–30:	27.8%	46.7%	45.0%
BMI 30–35:	33.3%	21.2%	22.3%
BMI $>$35:	5.60%	17.4%	16.3%

Table 4.2: **Patient Characteristics of TKR Patients.**

The mean improvement in HRQoL and mean NRSS per KL grade is shown in table 4.3 on the next page for THR patients and table 4.4 (p. 65) for TKR patients. In THR, patients with severe radiographic OA had a larger improvement in Physical Functioning than patients with mild radiographic OA. The improvement in other domains of HRQoL and the mean NRSS was similar for THR patients of all severities of radiographic OA. In TKR, patients with severe radiographic OA had a larger improvement in Physical functioning than patients with mild radiographic OA. Additionally, patients with severe radiographic OA had a larger improvement in General Health, a larger improvement in the Physical Component Summary Scale and a higher NRSS than patients with mild radiographic OA. The crude probabilities of achieving a MCID in each dimension of HRQoL are presented in table 4.5 (p. 66) for THR patients and table 4.6 (p. 67) for TKR patients. In THR, the probability of achieving a relevant improvement in Physical Functioning was higher in patients with severe radiographic OA than in patients with mild radiographic OA. The probability of achieving a satisfactory outcome was also higher in patients with severe radiographic OA than in patients with mild radiographic OA. The probability of achieving a relevant improvement in other domains of HRQoL was similar for THR patients of all severities of radiographic OA. In TKR, the probability of achieving a relevant improvement in Physical Functioning was higher in patients with severe radiographic OA than in patients with mild radiographic OA. Additionally, the probability of achieving a relevant improvement in General Health and the probability of

	Kellgren Grade 0–2: Mean Improvement (95%CI)	Kellgren Grade 3+4: Mean Improvement (95%CI)	Grade 0–2 vs 3+4: Mean Adjusted Difference (95%CI)	P-value
Physical Functioning	19.2 (14.2 – 24.1)	26.2 (22.4 – 30.0)	8.93 (2.14 – 15.7)	0.01
Role-Physical	36.3 (26.7 – 45.9)	42.2 (35.4 – 48.9)	6.39 (-5.89 – 18.7)	0.31
Bodily Pain	35.9 (30.4 – 41.3)	36.5 (32.8 – 40.2)	0.88 (-6.08 – 7.84)	0.80
General Health	0.60 (-3.50 – 4.60)	-1.50 (-4.50 – 1.50)	-0.66 (-5.66 – 4.34)	0.79
Vitality	9.30 (5.00 – 13.5)	3.70 (0.80 – 6.70)	-3.53 (-9.03 – 1.97)	0.21
Social Functioning	19.4 (13.6 – 25.2)	14.6 (10.7 – 18.4)	-4.11 (-11.2 – 2.97)	0.25
Role Emotional	6.90 (-1.10 – 14.9)	11.3 (4.70 – 17.8)	3.11 (-8.22 – 14.4)	0.59
Mental Health	7.20 (4.00 – 10.5)	4.60 (2.10 – 7.10)	-1.80 (-6.13 – 2.50)	0.41
PCS	10.7 (8.70 – 12.6)	11.2 (9.90 – 12.6)	1.94 (-0.57 – 4.44)	0.13
MCS	1.50 (-0.40 – 3.40)	-0.50 (-1.80 – 0.90)	-2.03 (-4.46 – 0.39)	0.10
NRS Satisfaction	8.5 (8.0 – 8.9)	8.9 (8.6 – 9.2)	0.3 (-0.2 – 0.9)	0.19

Table 4.3: **Improvement in Health-Related Quality of Life and Satisfaction after Hip Replacement: A Comparison Between Patients with Mild to Moderate and Severe Radiographical Pre-Operative Osteoarthritis.** Positive values indicate a higher mean improvement in HRQoL after THR in patients with Kellgren Grade 3+4, compared to Grade 0–2. The mean differences between radiographic severity are adjusted for age, sex, Charnley Comorbidity Classification and BMI and stratified for quartiles of follow-up.

achieving a satisfactory outcome was also higher in patients with severe radiographic OA than in patients with mild radiographic OA.

Discussion

At the population level, patients with severe radiographic OA improve more in Physical Functioning than patients with mild radiographic OA, both for THR and TKR. At the individual level, THR and TKR patients with severe radiographic OA have a larger probability of a relevant improvement in Physical Functioning than patients with mild radiographic OA. The effects of the preoperative severity of radiographic OA on Physical Functioning are more pronounced in TKR patients than in THR patients. Other domains of HRQoL do not appear to be influenced by the preoperative severity of OA, except General Health and the Physical Component Summary Scale in TKR patients. Additionally, patient satisfaction appears to be better in patients with more severe preoperative radiographic OA.

	Kellgren Grade 0–2: Mean Improvement (95%CI)	Kellgren Grade 3+4: Mean Improvement (95%CI)	Grade 0–2 vs 3+4: Mean Adjusted Difference (95%CI)	P-value
Physical Functioning	-2.10 (-10.5 – 6.30)	15.1 (11.7 – 18.5)	19.1 (8.48 – 29.7)	$<$0.001
Role-Physical	9.10 (-11.9 – 30.1)	20.6 (-13.5 – 27.7)	17.4 (-6.32 – 41.1)	0.15
Bodily Pain	14.5 (3.50 – 25.5)	25.2 (21.5 – 29.0)	9.02 (-3.43 – 21.5)	0.15
General Health	-9.10 (-16.9 – -1.30)	-1.50 (-3.80 – 0.80)	9.23 (1.31 – 17.2)	0.02
Vitality	-5.40 (-13.0 – 2.30)	1.20 (-1.40 – 3.80)	8.44 (-0.28 – 17.2)	0.06
Social Functioning	2.80 (-8.00 – 13.6)	8.90 (5.40 – 12.4)	7.44 (-4.18 – 19.1)	0.21
Role Emotional	4.50 (-17.5 – 26.6)	5.80 (-0.60 – 12.1)	8.87 (-11.8 – 29.6)	0.40
Mental Health	3.40 (-4.00 – 10.8)	3.00 (0.80 – 5.10)	0.29 (-6.93 – 7.50)	0.94
PCS	1.50 (-2.90 – 6.00)	6.40 (5.10 – 7.70)	5.64 (1.26 – 10.0)	0.01
MCS	0.10 (-4.30 – 4.40)	-0.30 (-1.60 – 1.00)	-0.18 (-4.45 – 4.10)	0.94
NRS Satisfaction	7.4 (6.1 – 8.6)	8.2 (7.9 – 8.6)	1.2 (0.1 – 2.4)	0.04

Table 4.4: **Improvement in Health-Related Quality of Life and Satisfaction after Knee Replacement: A Comparison Between Patients with Mild to Moderate and Severe Radiographical Pre-Operative Osteoarthritis.** Positive values indicate a higher mean improvement in HRQoL after TKR in patients with Kellgren Grade 3+4, compared to Grade 0–2. The mean differences between radiographic severity are adjusted for age, sex, Charnley Comorbidity Classification and BMI and stratified for quartiles of follow-up.

Limitations of the study include the participation rate and range of follow-up period after joint replacement. Although participation rates of 100% are feasible in small-scaled studies with hard endpoints,[34, 116] participation rates in epidemiological studies have been steadily declining in the last 30 years.[144] Even sharper declines have been reported in the past few years.[145] Unfortunately, the participation rate of this study follows this general trend, resulting in a participation rate of 46%. Therefore, we cannot exclude the presence of self-selection bias. In order to limit the extent of this bias, we have sent multiple reminders and have called all patients who did not answer our reminders and who did not return the questionnaire. As incentives, we have included an appealing information brochure in which the primary goals of the follow-up study were explained and a study pen as a small gift. Additionally, patients were urged to participate by their treating physician. However, the participation rate alone does not determine the extent of bias present in any particular study.[145] The difference between participants and non-participants is far more important.[146] As the found differences in demographics were of little clinical relevance, it is unlikely that the study results will be

	Kellgren Grade 0–2: Probability of Achieving MCID	Kellgren Grade 3+4: Probability of Achieving MCID	Grade 0–2 vs 3+4: Adjusted Odds Ratio (95%CI)	P-value
Physical Functioning	64 / 92: 69.6%	146 / 185: 78.9%	1.87 (0.97 – 3.60)	0.06
Role-Physical	55 / 92: 59.8%	124 / 185: 67.0%	1.50 (0.82 – 2.72)	0.19
Bodily Pain	71 / 92: 77.2%	141 / 185: 76.2%	1.03 (0.52 – 2.05)	0.93
General Health	62 / 92: 67.4%	117 / 185: 63.2%	0.91 (0.47 – 1.77)	0.78
Vitality	34 / 92: 37.0%	54 / 185: 29.2%	0.84 (0.46 – 1.55)	0.58
Social Functioning	42 / 92: 45.7%	80 / 185: 43.2%	0.87 (0.49 – 1.55)	0.64
Role Emotional	21 / 92: 22.8%	51 / 185: 27.6%	1.01 (0.51 – 2.01)	0.98
Mental Health	17 / 92: 18.5%	40 / 185: 21.6%	1.26 (0.62 – 2.58)	0.53
NRS Satisfaction $>$ 8	53 / 92: 57.6%	136 / 185: 73.5%	1.95 (1.06 – 3.59)	0.03

Table 4.5: **Improvement in Health-Related Quality of Life and Satisfaction after Hip Replacement: A Comparison Between Patients with Mild to Moderate and Severe Radiographical Pre-Operative Osteoarthritis.** Odds Ratios > 1 indicate a higher probability of achieving a Minimal Clinically Important Difference in HRQoL after THR in patients with Kellgren Grade 3+4, compared to Grade 0–2. The odds ratios are adjusted for age, sex, Charnley Comorbidity Classification and BMI and stratified for quartiles of follow-up.

severely biased. Finally, the patient demographics of our study population were similar to those of large-scaled national joint registry studies, regarding age, gender, Charnley classification and BMI.[143, 175]

The follow-up period after joint replacement varies between 2 and 5 years. Although a residual effect of follow-up length cannot be excluded, we do not think this is very plausible, as recent evidence suggests that the improvement in HRQoL is sustained up to 5 years after joint replacement surgery.[147, 148]

Although joint replacements are highly effective in improving HRQoL at the group level,[1] this is not the case for each individual patient, judging from the relatively high dissatisfaction rates.[19, 20] Studying HRQoL at the individual level, using the probability of achieving a clinically important difference as an outcome measure, enables a better prediction of a successful outcome. Moreover, it could provide a helpful way to fine-tune the indication for joint replacement, for which there are no clear cut-off points currently available.[176]

	Kellgren Grade 0–2: Probability of Achieving MCID	Kellgren Grade 3+4: Probability of Achieving MCID	Grade 0–2 vs 3+4: Adjusted Odds Ratio (95%CI)	P-value
Physical Functioning	5 / 22: 22.7%	105 / 191: 55.0%	5.44 (1.45 – 20.3)	0.01
Role-Physical	9 / 22: 40.9%	88 / 191: 46.1%	1.46 (0.49 – 4.32)	0.50
Bodily Pain	15 / 22: 68.2%	136 / 191: 71.2%	1.15 (0.32 – 4.16)	0.83
General Health	9 / 22: 40.9%	122 / 191: 63.9%	3.56 (1.23 – 10.4)	0.02
Vitality	8 / 22: 36.4%	86 / 191: 45.0%	1.09 (0.35 – 3.44)	0.88
Social Functioning	7 / 22: 31.8%	98 / 191: 51.3%	2.84 (0.87 – 9.32)	0.08
Role Emotional	6 / 22: 27.3%	41 / 191: 21.5%	0.85 (0.26 – 3.02)	0.85
Mental Health	8 / 22: 36.4%	79 / 191: 41.4%	2.79 (0.70 – 11.2)	0.15
NRS Satisfaction >8	9 / 22: 40.9%	116 / 191: 60.7%	2.25 (0.78 – 6.52)	0.14

Table 4.6: **Improvement in Health-Related Quality of Life and Satisfaction after Knee Replacement: A Comparison Between Patients with Mild to Moderate and Severe Radiographical Pre-Operative Osteoarthritis.** Odds Ratios > 1 indicate a higher probability of achieving a Minimal Clinically Important Difference in HRQoL after TKR in patients with Kellgren Grade 3+4, compared to Grade 0–2. The odds ratios are adjusted for age, sex, Charnley Comorbidity Classification and BMI and stratified for quartiles of follow-up.

Regardless of age, gender, co-morbidity and BMI, we have shown that joint replacement patients with severe preoperative OA have a better prognosis in improvement in Physical Functioning and patient satisfaction with the surgical results. These effects are more pronounced in TKR patients than in THR patients, which might be explained in part by biomechanical factors. The hip joint is a relatively simple ball and socket joint, which is adequately mimicked by a THR. The biomechanical aspects of the knee joint are more difficult to imitate, as the knee is a pivotal hinge joint with 6 degrees of freedom. These degrees of freedom are generally not restored after TKR, which is substantiated in kinematic and kinetic studies.[163] This additional disadvantage of TKR patients who underwent joint replacement for mild radiographic OA is reflected in a smaller increase in Physical Functioning than THR patients who underwent joint replacement for mild radiographic OA. Additionally, the odds of achieving a MCID in Physical Functioning is smaller and the difference in satisfaction is larger.

Clinically, these are promising findings, as dissatisfaction rates are higher in TKR patients than in THR patients.[13, 15] Patient satisfaction is thought to be closely related to unfulfilled expectations. Although patient expectations of THR and TKR are similar,

recent evidence suggests that THR meets important patient expectations better than TKR.[15, 166, 177] Our findings could lead to a more fitting expectation management regarding the expected improvement in Physical Functioning, using a single predictor. This improvement in expectation management might lead to higher satisfaction rates.

Plain radiographs have a number of appealing aspects. In the first place, they are inexpensive and easily available, as they are currently a part of the clinical work-up to joint replacement. Secondly, due to the non-invasive character of the test, radiographs are a patient-friendly modality. Finally, they offer a more objective approach to joint complaints. These aspects would make it easy to implement the KL grade in clinical practice, in order to predict HRQoL and satisfaction after joint replacement.

Part III

Research Methodology

5

Revision surgery is Overestimated in Hip Replacement

JC Keurentjes[1], M Fiocco[2], BW Schreurs[3], BG Pijls[1], KA Nouta[1], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Orthopaedic Surgery, Radboud University Nijmegen Medical Center.

Bone & Joint Research. 1 (10), 258-262.

Abstract

Introduction The Kaplan-Meier estimation is widely used in orthopedics to calculate the probability of revision surgery. Using data from a long-term follow-up study, we aimed to assess the amount of bias introduced by the Kaplan-Meier estimator in a competing risk setting.

Methods We describe both the Kaplan-Meier estimator and the competing risk model, and explain why the competing risk model is a more appropriate approach to estimate the probability of revision surgery when patients die in a hip revision surgery cohort. In our study, a total of 62 acetabular revisions were performed. After a mean of 25 years, no patients were lost to follow-up, 13 patients had undergone revision surgery and 33 patients died of causes unrelated to their hip.

Results The Kaplan-Meier estimator overestimates the probability of revision surgery in our example by 3%, 11%, 28%, 32% and 60% at five, ten, 15, 20 and 25 years, respectively. As the cumulative incidence of the competing event increases over time, as does the amount of bias.

Discussion Ignoring competing risks leads to biased estimations of the probability of revision surgery. In order to guide choosing the appropriate statistical analysis in future clinical studies, we propose a flowchart.

Introduction

One of the most important outcome measures in orthopaedic surgery is the time to a certain event. In joint replacement surgery, for instance, the time to revision surgery is seen as the most important determinant of the clinical success of any prosthesis. Techniques from the field of survival analysis, such as the Kaplan-Meier estimator,[22] have been used to estimate time to revision surgery since the 1980s.[178, 179] The time from implantation of a prosthesis until a specified event of interest is used in survival analyses. An important advantage of survival analyses is that these techniques allow analyses with “censored data”, i.e. data concerning patients for which revision surgery has not yet taken place within the study period.[22] If the endpoint of interest has not

yet occurred at the end of the observation window, the event time is censored. The probability of revision surgery can be estimated with the Kaplan-Meier estimator at any specific point in time.

At first glance, the Kaplan-Meier estimator seems ideal for orthopaedics since analyses can be performed before revision surgery has occurred in all patients. However, this method makes a number of assumptions.[115, 180] The Kaplan-Meier estimator is specifically developed for studies with a single time to a certain event, which in turn is able to be censored. The assumption of independence of the time to event and the censoring distributions is of critical importance. The probability of the event of interest is estimated by assuming that patients whose time is censored have the same probability of revision at any later time. When estimating the time to revision surgery, often more types of events play a role, which may prevent the event of interest from occurring. For instance, revision of an implant may be unobservable because the patient dies. In this particular case, death is a competing event, which poses a competing risk — a risk that may be high, especially in studies with long-term follow-up.

The Kaplan-Meier method of censoring patients who experience a competing event is not ideal when the estimation of the probability of the event of interest is the goal, since this implicitly assumes that the event of interest still could occur after the time point at which censoring occurred.[181–183] If a patient does experience a competing event, the event of interest can no longer occur: therefore the potential contribution to the estimate from this patient should become zero. The probability of the event of interest must be estimated by taking into account the probability of the competing events; ignoring the competing risks leads to a biased estimation of the probability of the event of interest (see Appendix 1, available at *http://www.bjr.boneandjoint.org.uk/content/1/10/258/suppl/DC1*).[180, 184–186]

In this study we compare the Kaplan-Meier estimator with the cumulative incidence estimator in a competing risk setting and show how the level of bias introduced by violating critical assumptions of the Kaplan-Meier estimator. We propose a simple

algorithm to help select the appropriate data analysis technique to estimate the probability of revision surgery in future studies. In order to illustrate these statistical methods, developed by Kaplan and Meier[22] and Bernoulli,[185, 186] we used data from a previous cohort of acetabular revision patients.[34]

Methods

In our published cohort study, 62 acetabular revisions were performed in 58 patients between January 1979 and March 1986, at the Radboud University Medical Center in Nijmegen, The Netherlands.[34] There were 13 men and 45 women with a mean age at revision of 59.2 years (range: 23 – 82). Revision was undertaken using impacted morsellised bone grafts and a cemented acetabular component in all cases. All patients were followed prospectively with yearly clinical and radiological assessments.

Competing risks versus Kaplan-Meier Competing risks are applied to situations where more than one competing endpoints are possible. Their competing in that one event will preclude the other occurring. In our situation there are two different endpoints: revision surgery and death. The occurrence of death prevents the occurrence of the event of interest, namely revision surgery. The competing risks model can be represented as an initial state (alive after initial revision surgery) and two different competing endpoints: revision surgery and death. We are interested in the probability of revision surgery (event of interest) in the presence of the competing event of death — which clearly prevents the occurrence of revision.

The Kaplan-Meier estimator is often used to estimate this probability. However, in this model the competing cause endpoints (i.e., death) are treated as censored observations. If a patient has experienced death, he or she has zero probability of experiencing the event of interest, and this must be considered in the model.

The cumulative incidence estimator is used to estimate the probability of each competing event. The *cumulative incidence function of cause k* is defined as the probability of failing from *cause k* before *time t*. Here we are interested in the cumulative incidence function of revision surgery in the presence of death.

Statistical analysis All analyses concerning competing risks models have been performed using the mstate library[187, 188] in R.[43] For technical details concerning the software, see de Wreede et al.[187, 188]

Results

At a mean of 23 years (20 to 25) after surgery, no patients were lost to follow-up. A total of 13 hips in 12 patients had undergone revision surgery, and 30 patients (33 hips) had died of causes unrelated to their hip surgery (table 5.1).

The estimated survival rates with revision surgery as the endpoint obtained by applying the Kaplan-Meier method at five, ten, 15, 20 and 25 years were, respectively, 98% (95% confidence interval (CI): 95 – 100), 93% (95% CI: 86 – 99), 81% (95% CI: 67 – 95), 75% (95% CI: 57 – 93) and 66% (95% CI: 49 – 83).

The estimated risk of revision surgery (1 - estimated survival of the implant) obtained with the Kaplan-Meier estimator, is shown in figure 5.1 on the following page. These estimated risks of revision surgery were therefore 2%, 7%, 19%, 25% and 34% at five, ten, 15, 20 and 25 years, respectively. The cumulative incidence estimators for both competing events, i.e. revision surgery and death, are shown in figure 5.2 (p. 77). The cumulative incidence estimator of revision surgery by the competing risks method at five, ten, 15, 20 and 25 years is 2%, 6%, 15%, 18% and 21%, respectively. The cumulative incidence of death represents the probability of dying before revision surgery. If death occurs first, the observation will not be considered censored in the competing

	Acetabular revisions	Patients
Total	62	58
Lost to follow-up	0	0
Died without further surgery	33	30
Revisions	13	12
Septic loosening	2	
Aseptic loosening	8	
Mismatch during femoral revision	1	
Wear	2	

Table 5.1: **Details of the 62 consecutive acetabular revisions.**

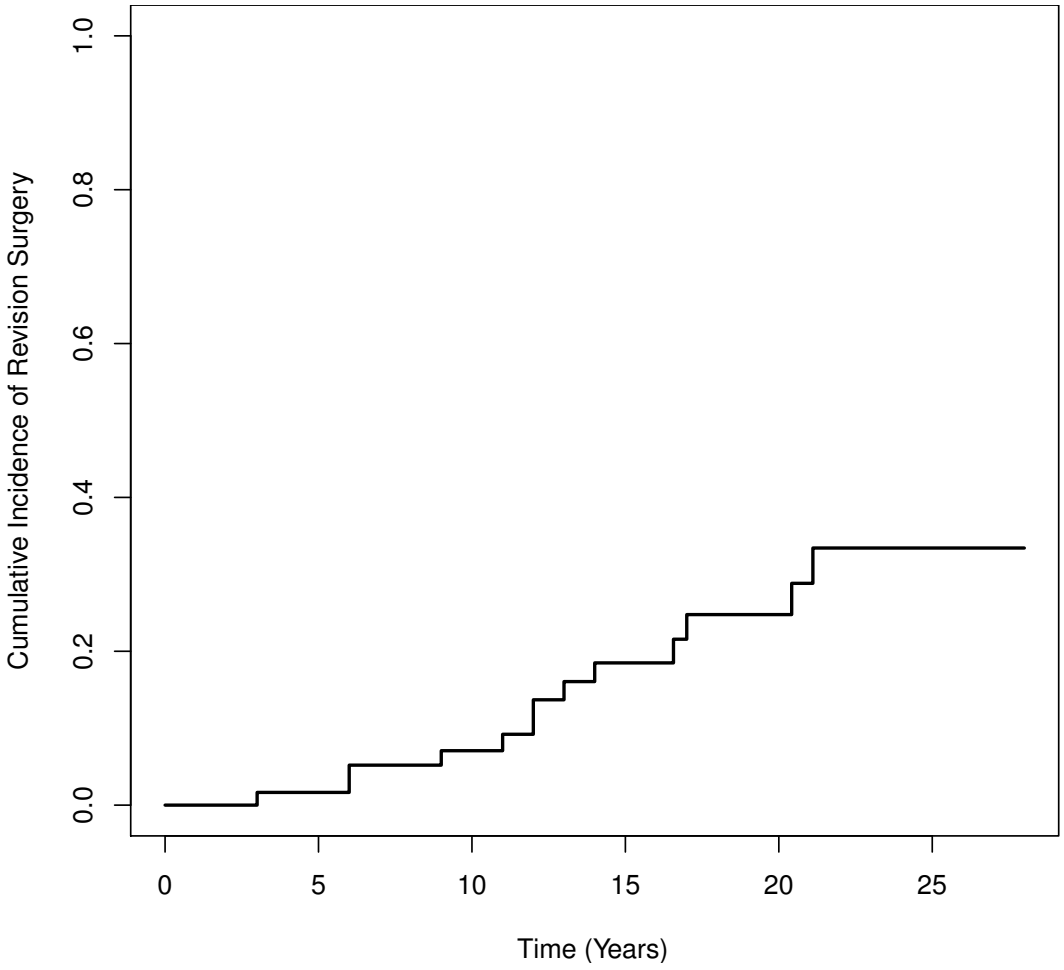


Figure 5.1: **Kaplan-Meier curve showing the cumulative incidence of revision surgery.** The risk of revision surgery in the Kaplan-Meier approach can be represented as: risk at time t = 1 - survival at time t.

risk approach (in contrast to the Kaplan-Meier approach), but it will contribute to the competing event of death. In the dataset described above, the Kaplan-Meier model can be seen to overestimate the probability of revision surgery by 3%, 11%, 28%, 32% and 60% at five, ten, 15, 20 and 25 years, respectively (fig. 5.3 (p. 78)).

Discussion

In the current orthopaedic literature, the Kaplan-Meier estimator is an accepted standard in estimating the probability of revision surgery in cohort studies of any type of joint replacement. In the absence of competing risks, this method is valid. However, in the presence of competing risks, the Kaplan-Meier estimator overestimates the probability of revision surgery. In our example, the probability of revision surgery is overestimated

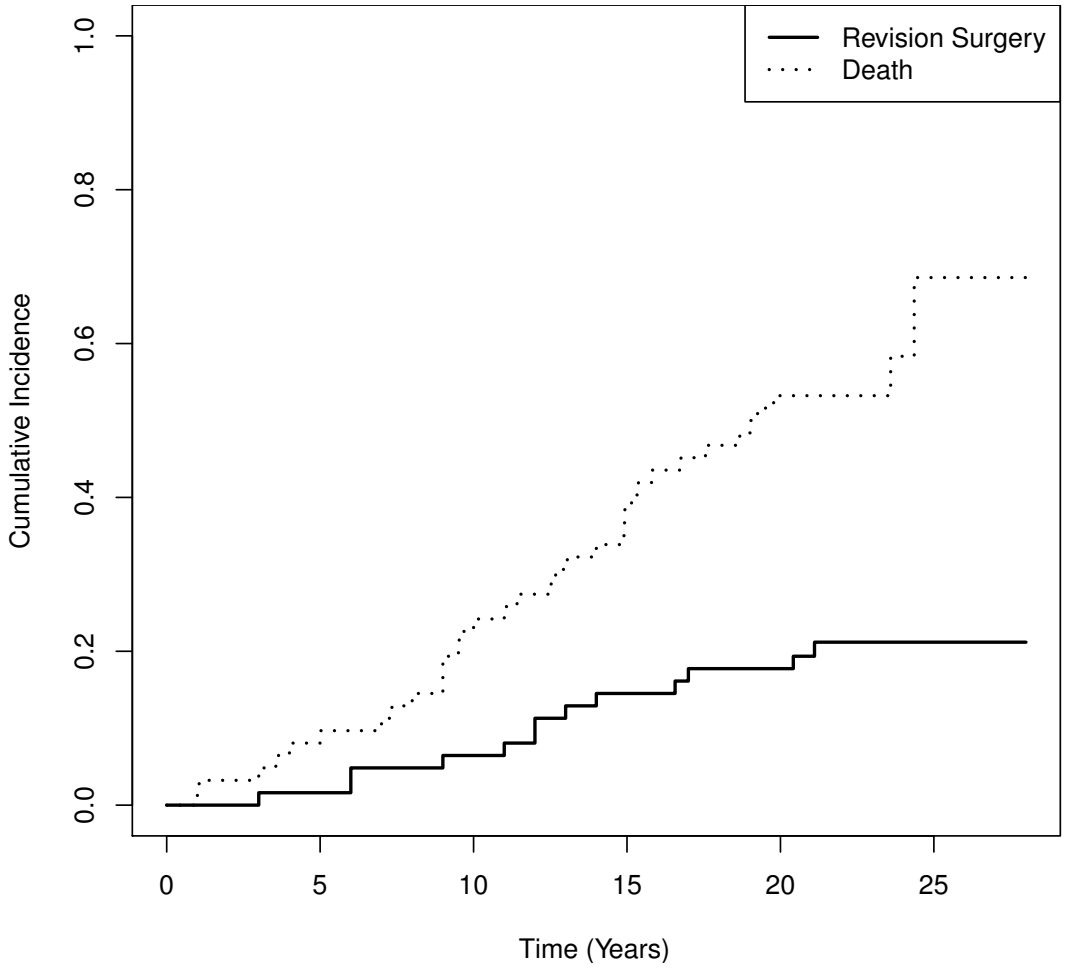
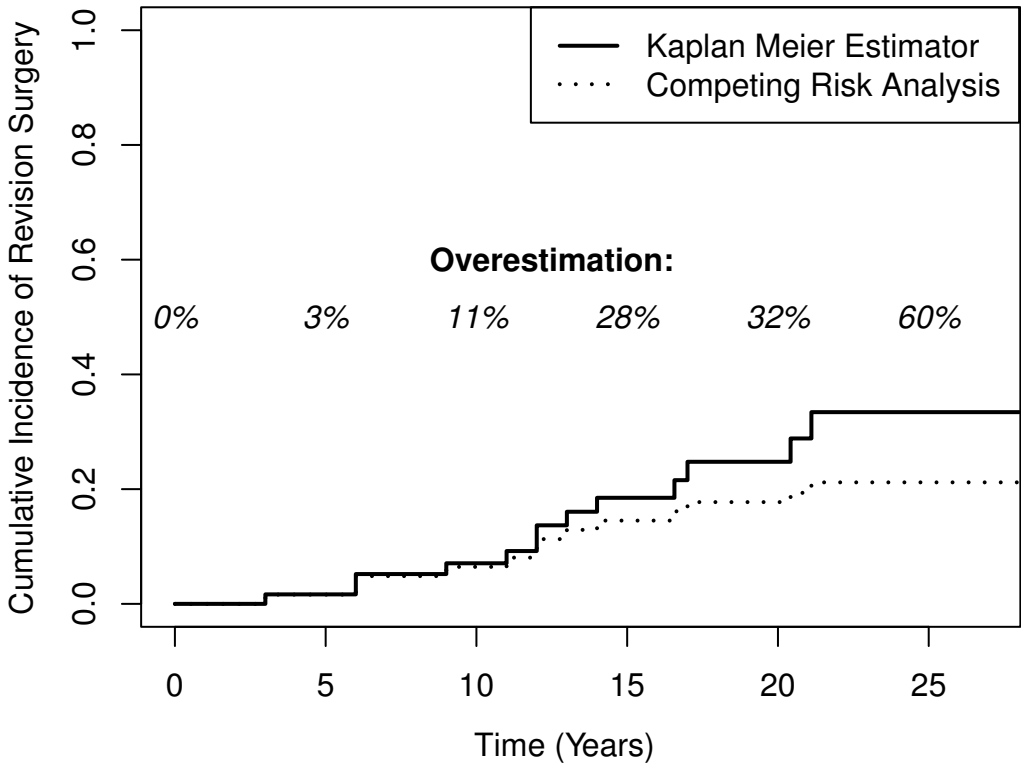


Figure 5.2: **Kaplan-Meier curve showing the cumulative incidence of revision surgery.** The risk of revision surgery in the Kaplan-Meier approach can be represented as: risk at time t = 1 - survival at time t.

by 60% at a follow-up of 25 years. In the Kaplan-Meier approach failures from the competing causes are treated as censored observations. Individuals who will never be revised because they have died, are censored and thus treated as if they still could be revised. In other words, the Kaplan-Meier estimator allows patients to be revised after they have died. Clearly, this results in an incorrect or biased estimate of the actual probability of revision surgery at that specific time point.

When competing risks are absent (i.e., the competing event death has not occurred), the Kaplan-Meier estimator gives a valid estimation of the probability of revision surgery. However, in our example involving a long follow-up, competing events such as death do occur frequently. Also, it can be seen from our dataset that the first patient died



N Dead	0	6	15	24	33	35
N Revised	0	1	4	9	11	13
N At Risk	62	55	43	29	18	14

Figure 5.3: **Comparison of cumulative incidence of revision surgery estimated with the Kaplan-Meier estimator and the competing risks method.** The discrepancy between the lines represents the bias, which is introduced by erroneous usage of the Kaplan-Meier estimator.

as early as one year after surgery (fig. 5.2 on the previous page). By five years after the initial surgery, a total of six patients had died, compared with only one patient who had undergone revision surgery, resulting in a 3% overestimation of the probability of revision surgery (fig. 5.3). In other words, the hazard of the competing events is considerable, leading to an overestimation of the revision surgery probability, even at mid-term follow-up.

In this paper a competing risks model has been applied to a cohort where only two competing events are present. However, in other clinical situations, more competing events can occur. Consider estimating the probability of revision surgery due to a specific

event, for instance the probability of revision surgery due to recurrent dislocations. In this situation, there are three competing events: revision surgery for recurrent dislocations, revision surgery for any other reason and death of a patient. The competing risk model can easily be extended to deal with another competing event.

From a statistical point of view, competing risk analysis should be used whenever competing risks are present. In order to aid in deciding which analysis should be used to estimate the probability of revision surgery in future clinical studies, we propose a simple algorithm (fig. 5.4 on the next page). Every clinical study that investigates the probability of revision surgery should address the occurrence of competing events. When no competing events have occurred, the Kaplan-Meier estimator of revision surgery will be valid. However, whenever any competing event occurs, the Kaplan-Meier estimator will introduce bias. The resulting bias is greater when the "competition" is heavier, i.e. when the hazard of the competing events is larger. See Appendix 2 of the Supplementary Material for a concise summary of necessary variables to perform a competing risk analysis (available at *http://www.bjr.boneandjoint.org.uk/content/1/10/258/suppl/DC1*).

Recently, minimal clinically important differences (MCIDs) have gained attention in the literature.[141, 169, 170] Using MCIDs, patients can be classified as responders or non-responders to a particular therapy. Theoretically, one could investigate the time to a MCID after joint replacement, using MCIDs in health-related quality of life (HRQoL). However, contrary to the occurrence of revision surgery or the first occurrence of a complication,[189] which can be assessed over a time period, whether or not a patient has attained a MCID in HRQoL is typically measured using a questionnaire at a specific point in time. Neither the Kaplan-Meier estimator nor a competing risk model is an appropriate approach, unless the assessment of the occurrence of an MCID is repeated at small time intervals.

The competing risk analysis can be performed using the mstate library[187, 188] in R.[43] R and the mstate package are both freely available at The R Project for Statistical Computing and The Comprehensive R Archive Network.

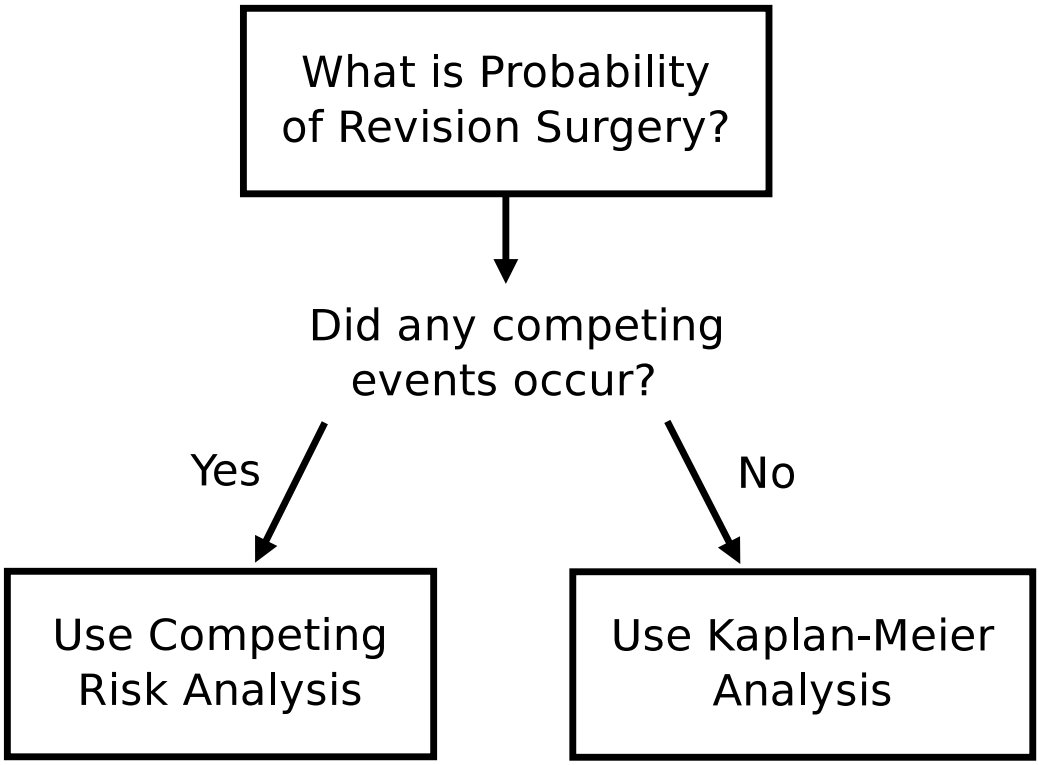


Figure 5.4: **Algorithm detailing the appropriate data analysis technique to estimate the probability of revision surgery.** The possibility and actual occurrence of competing events should be assessed in order to determine the appropriate data analysis technique.

6

Minimal Clinically Important Differences in Health-Related Quality of Life after Total Hip or Knee Replacement

JC Keurentjes[1], FR Van Tol[1], M Fiocco[2], JW Schoones[3], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Walaeus Library, Leiden University Medical Center.

Bone & Joint Research. 1 (5), 71-77.

Abstract

Objectives We aimed first to summarise minimal clinically important differences (MCIDs) after total hip (THR) or knee replacement (TKR) in health-related quality of life (HRQoL), measured using the Short-Form 36 (SF-36). Secondly, we aimed to improve the precision of MCID estimates by means of meta-analysis.

Methods We conducted a systematic review of English and non-English articles using MEDLINE, the Cochrane Controlled Trials Register (1960 – 2011), EMBASE (1991 – 2011), Web of Science, Academic Search Premier and Science Direct. Bibliographies of included studies were searched in order to find additional studies. Search terms included MCID or minimal clinically important change, THR or TKR and Short-Form 36. We included longitudinal studies that estimated MCID of SF-36 after THR or TKR.

Results Three studies met our inclusion criteria, describing a distinct study population: primary THR, primary TKR and revision THR. No synthesis of study results can be given.

Conclusions Although we found MCIDs in HRQoL after THR or TKR have limited precision and are not validated using external criteria, these are still the best known estimates of MCIDs in HRQoL after THR and TKR to date. We therefore advise these MCIDs to be used as absolute thresholds, but with caution.

Introduction

Total hip (THR) and knee replacement (TKR) are effective surgical interventions, which alleviate pain and improve function and health-related quality of life (HRQoL) in patients with end-stage degeneration of the hip or knee joint, respectively.[190] Typically, studies report the mean improvement in HRQoL at the population level, which provides information for the average patient in a population. However, this information may not be meaningful for individual patients encountered in clinical practice, who will be concerned with the likelihood that they will experience a meaningful improvement in return for the risk undertaken when undergoing an intervention.[191] More relevant for the individual patient therefore is the minimal clinically important difference (MCID), defined as the

minimal difference in scores of an outcome measure that is perceived by patients as beneficial or harmful.[167, 168] The MCID enables patients to be classified as either a responder or a non-responder to a particular therapy, based on their own assessment of their pre- and post-operative HRQoL. Additionally, the MCID allows an estimation of the probability of a relevant improvement in HRQoL of a particular therapy.

Expected benefits of treatment must be weighed against its adverse effects, inconvenience and costs.[192] Therefore, there is not necessarily a single MCID value for any one outcome measure of HRQoL, which can be used for all applications and patient samples.[193] For instance, the benefits of treatment in patients suffering from end-stage osteoarthritis are considerably larger for THR and TKR compared with rehabilitational interventions. On the other hand, the risk of adverse effects is also considerably higher. These differences complicate the direct use of MCIDs in HRQoL as established for rehabilitational interventions,[194] in THR or TKR patients. The use of specific MCIDs in HRQoL after THR or TKR should be encouraged.

MCIDs can be established using two different methods. Anchor-based approaches use an external indicator to assign patients into several groups reflecting different amounts of change in health status.[193] The within-person global change rating is often used as an anchor, which is measured using Likert scales, ranging from five to 15 options.[193] Positive MCIDs are usually estimated by the mean difference between pre- and post-operative scores of patients, who indicate that their condition is “somewhat better”; negative MCIDs are usually estimated by the mean difference between pre- and post-operative scores of patients who indicate that their condition is “somewhat worse”.[192, 195] Distribution-based methods offer another approach in the estimation of MCIDs, which interpret results in terms of the relation between the magnitude of effect and some measure of variability in results.[192] Individual effect size standards are often used to estimate the MCID, which is defined as the difference between a patient’s pre- and post-operative HRQoL scores, normed to the standard deviation of the pre-operative scores.[195] Generally accepted individual effect size standards are

equal to the group effect size standards, as defined by Cohen.[196] Therefore, the MCID is calculated by multiplying the standard deviation of patients at baseline by 0.5.

Recently Quintana et al[169] and Escobar et al[170] have estimated MCIDs for the SF-36 after THR and TKR. However, these authors have advised against using the found estimates of MCIDs as absolute thresholds, due to the imprecision of these estimates caused by small sample sizes. The precision of an estimate can be enhanced by pooling results of multiple studies in a meta-analysis. Therefore, the purpose of our study was to enhance the precision of the MCIDs after THR and TKR, by means of a systematic review and meta-analysis.

Methods

This systematic review was performed in November 2011, using the PRISMA-Statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) as a guideline in the development of the study protocol and the report of the current study.[36] The inclusion criteria and methods of analysis were specified in advance and documented in a protocol.

Information sources and search strategy Longitudinal studies that estimate the MCID in HRQol, measured using the Short-Form 36 (SF-36), after primary or revision THR or TKR, were eligible for inclusion. No language, publication date, or publication status restrictions were imposed. Studies were identified by searching electronic databases. No limits were applied for language and foreign papers were translated. This search strategy was applied to PubMed, MEDLINE, Embase, Web of Science, COCHRANE, ScienceDirect and Academic Search Premier. The search was run on 8 November 2011. The following search terms were used in PubMed, and were adapted for the other databases:
(Mcid[tw] OR cid[tw] OR “Minimal clinically important differences” OR “Minimal clinically important difference” OR “clinically important differences” OR “clinically important difference” OR MCIC[tw] OR “Minimal clinically important changes” OR “Minimal

clinically important change" OR "clinically important changes" OR "clinically important change" OR "Minimal clinical important differences" OR "Minimal clinical important difference" OR "clinical important differences" OR "clinical important difference" OR "clinical important changes" OR "minimal detectable change" OR "minimal detectable changes" OR "minimally detectable change" OR "meaningful changes" OR "meaningful change") AND (tka[tw] OR "knee replacement arthroplasty" OR "knee arthroplasty" OR "knee replacement" OR "knee prosthesis" OR tha[tw] OR "hip replacement arthroplasty" OR "hip arthroplasty" OR "hip replacement" OR "hip prosthesis" OR "knee" OR "knees" OR "hip" OR "hips") AND ("SF36" OR "SF-36" OR "short form 36" OR "shortform 36")

Study selection Two authors (JCK and FRvT) independently screened titles and abstracts of the papers resulting from the database search using predefined eligibility criteria. Papers were considered eligible for inclusion if they met two criteria; they were to concern primary or revision THR or TKR and should include an estimate of a MCID. The full text of all included papers, based on titles and abstracts, were screened using the same inclusion criteria. Disagreements between reviewers were resolved by consensus.

HRQoL measured using SF-36 The SF-36 consists of 36 items, covering eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health), for which a transformed score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms).[197]

Data collection process and data items Both authors extracted the data independently, using a predefined data extraction form. Areas of disagreement or uncertainty were resolved by consensus. Estimates of MCIDs were extracted from included studies. For anchor-based estimates of MCIDs, we extracted the number of patients, on which the estimate was based, and the standard deviation. For distribution-based estimates, we extracted the number of patient on which the estimate was based. Additionally, study characteristics, concerning follow-up period, sample size, proportion of patients who underwent joint replacement for osteoarthritis, proportion of males,

mean patient age and proportion lost to follow-up, were collected.

Risk of bias in individual studies We assessed the risk of bias in the included studies through a modified Newcastle-Ottawa Quality Assessment Scale,[198] which included the following questions: “which approach was used to estimate the MCID?” (anchor-based versus distribution-based); “was any form of additional validation performed?” (yes / no); “was the study population representative of THR or TKR in general?” (truly representative / somewhat representative / selected population / not enough information given); “was the follow-up adequate?” (no loss to follow-up / < 5% lost to follow-up (unlikely to bias results) / > 5% lost to follow-up (results possibly biased)). We chose the cut-off point of 5% lost to follow-up according to Pijls et al,[37] who established this threshold for observational studies in orthopaedic literature, using a Delphi approach to form consensus between a group of experts in the fields of THR, TKR or evidence-based medicine.

Summary measures and planned methods of analysis The primary outcome measure was the MCID in HRQoL, measured using SF-36, for primary THR, primary TKR, revision THR and revision TKR. Whenever possible, estimates of MCIDs were pooled using inverse variance weighting. 95% confidence intervals (CI) were calculated for all MCID estimates.

Results

Study selection The search strategy revealed a total of 126 results (fig. 6.1 on the facing page). After removal of duplicate entries, 114 unique papers remained. Screening of titles and abstracts revealed 29 papers eligible for inclusion. Further assessment of eligibility, based on full-text papers, led to the exclusion of 26 papers: two did not address THR or TKR and 24 presented no estimation of an MCID. This left three papers, describing three studies, for further analysis.[169, 170, 199]

Study characteristics An overview of the study characteristics of all included studies is presented in table 6.1 (p. 88). Quintana et al[169] describe the MCID in SF-36 after

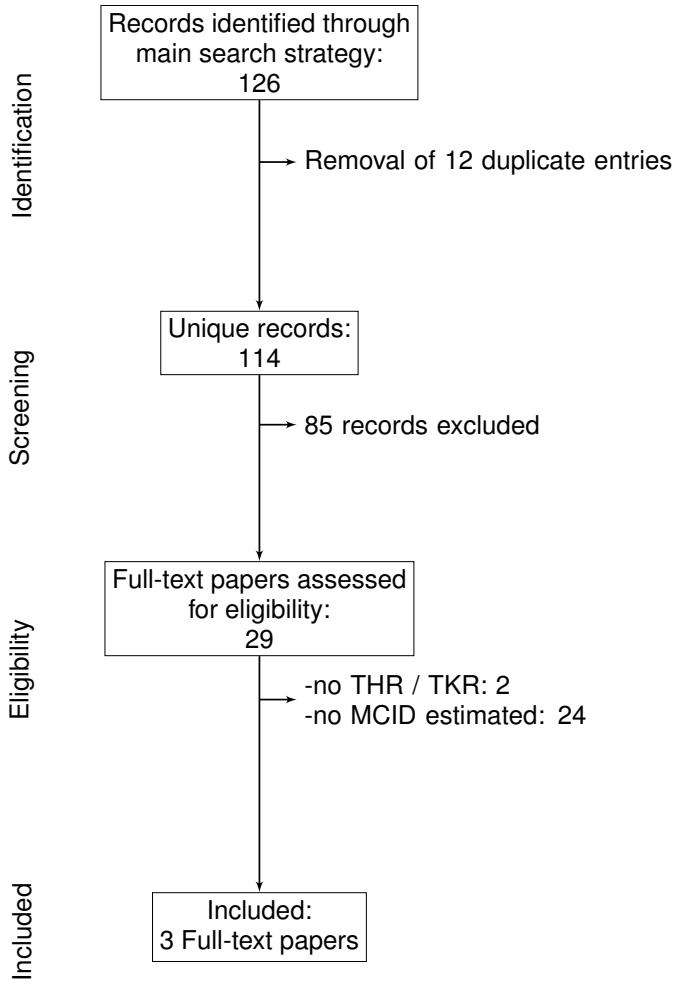


Figure 6.1: **Flow-chart of study inclusion.** THR, total hip replacement; TKR, total knee replacement; MCID, minimal clinically important difference.

primary THR at follow-up periods of six months and two years; Escobar et al[170] describe the MCID in SF-36 after primary TKR at follow-up periods of six months and two years; and Shi et al[199] describe the MCID in SF-36 after revision THR at a follow-up period of six months (table 6.1 on the following page). All included studies were multi-center studies. All studies estimated positive MCIDs (i.e. the minimal difference in scores of the SF-36 that is perceived by patients as beneficial); no study estimated negative MCIDs (the minimal difference in scores of an outcome measure that is perceived by patients as harmful). The sample for the estimation of the MCIDs was 43 patients after six months and 33 after two years for Quintana et al[169]; 76 after six months and 65 after two years for Escobar et al[170]; and 67 after six months for Shi et al[199] (table 6.1 on the next page). The indication for joint replacement was

Ref.	Intervention	Follow-up	Setting	+/-[1]	Study Sample Size	Effective Sample size[2]	OA[3] (%)	Male (%)	Mean age (yrs)	Lost to follow-up (%)
[169]	Primary THR	6 months	Multi-centre	+	485	43	100	49,3	69,4	21,86
	Primary THR	2 years	Multi-centre	+	310	33	100	49,3	69,4	36,08
[170]	Primary TKR	6 months	Multi-centre	+	423	76	100	25	71,6	22,39
	Primary TKR	2 years	Multi-centre	+	364	65	100	25	71,6	33,21
[199]	Revision THR	6 months	Multi-centre	+	67	67	n/a	56,7	70,2	16,25

Table 6.1: **Study characteristics of the three included studies.** 1: Positive or negative MCID estimated; 2: Sample size for MCID estimation; 3: Osteoarthritis.

Authors	Follow-up	MCID methodology	Additional validation	Representativeness of study population	Adequacy of follow-up
Quintana[169]	6 months	Anchor-based	No	Truly representative	>5% lost
	2 years	Anchor-based	No	Truly representative	>5% lost
Escobar[170]	6 months	Anchor-based	No	Truly representative	>5% lost
	2 years	Anchor-based	No	Truly representative	>5% lost
Shi[199]	6 months	Distribution-based	No	Description of cohort incomplete	>5% lost

Table 6.2: **Risk of bias within the three included studies.**

osteoarthritis in all patients of Quintana et al[169] and Escobar et al,[170] while Shi et al[199] offered no statement of the indication for joint replacement (table 6.1). In all studies, some patients were lost to follow-up.

Risk of bias within studies An overview of the risk of bias within studies is presented in table 6.2. Two studies used anchor-based approaches to estimate the MCID,[169, 170] while the other used a distribution-based approach.[199] No study performed any form of additional validation. The study populations of Quintana et al[169] and Escobar et al[170] are truly representative of THR and TKR patients in general, while Shi et al[199] did not provide enough information to assess the representativeness by leaving out the indication of joint replacement. All studies lost >5% of patients to follow-up, rendering a possibility of biased results.

Synthesis of results All studies have described a distinct study population, precluding any meaningful synthesis of study results. An overview of the results of all individual studies is presented in table 6.3 and figure 6.2 and figure 6.3. The MCIDs are presented with 95% confidence intervals for each of the SF-36 domains in primary

		MCID (95% confidence interval)	
	SF-36 domain	At six months	At two years
Primary THR[169]	Physical functioning	20.40 (14.4 – 26.4)	8.29 (-1.8 – 18.4)
	Role physical	10.78 (1.5 – 20.0)	11.00 (-1.3 – 23.3)
	Bodily pain	14.67 (6.8 – 22.6)	18.34 (9.1 – 27.6)
	General health	0.40 (-5.2 – 6.0)	-6.37 (-10.9 – -1.9)
	Vitality	10.14 (3.1 – 17.2)	14.51 (6.4 – 22.6)
	Social functioning	8.63 (0.9 – 16.4)	17.97 (7.8 – 28.1)
	Role emotional	-6.45 (-24.5 – 11.6)	20.83 (-0.6 – 42.3)
	Mental health	8.99 (2.3 – 15.7)	16.15 (9.0 – 23.3)
Primary TKR[170]	Physical functioning	11.57 (6.5 – 16.7)	11.07 (5.8 – 16.3)
	Role physical	11.69 (3.8 – 19.6)	13.16 (3.5 – 22.8)
	Bodily pain	16.86 (9.7 – 24.0)	6.69 (-0.4 – 13.8)
	General health	0.85 (-3.2 – 4.9)	-7.30 (-11.3 – -3.3)
	Vitality	3.86 (-1.7 – 9.4)	3.44 (-2.2 – 9.1)
	Social functioning	11.66 (3.7 – 19.6)	6.15 (-1.7 – 14.0)
	Role emotional	7.65 (-4.5 – 19.8)	2.42 (-9.2 – 14.1)
	Mental health	-0.32 (-5.5 – 4.9)	4.02 (-1.7 – 9.7)
Revision THR[199]	Physical functioning	3.25 (2.8 – 3.9)	-
	Role physical	4.78 (4.1 – 5.8)	-
	Bodily pain	14.91 (12.7 – 18.0)	-
	General health	14.12 (12.1 – 17.0)	-
	Vitality	22.81 (19.5 – 27.5)	-
	Social functioning	15.83 (13.5 – 19.1)	-
	Role emotional	19.98 (17.1 – 24.1)	-
	Mental health	12.37 (10.6 – 14.9)	-

Table 6.3: **Minimal clinically important differences (MCIDs) in Short-Form 36 (SF-36) domains after primary and revision total hip replacement (THR) and primary total knee replacement (TKR).**

TKR and primary and revision THR at six months (fig. 6.2 on the following page, table 6.3)[169, 170, 199] and for primary TKR and THR at two years post-operatively (fig. 6.3 (p. 91), table 6.3).[169, 170]

Discussion

We have found one study describing MCIDs in SF-36 after primary THR,[169] one after primary TKR[170] and one after revision THR[199]; we did not find any studies describing MCIDs after revision TKR. As all studies have described a distinct study population, no synthesis of study results can be given. Therefore, we were unable to improve the precision of each MCID estimate.

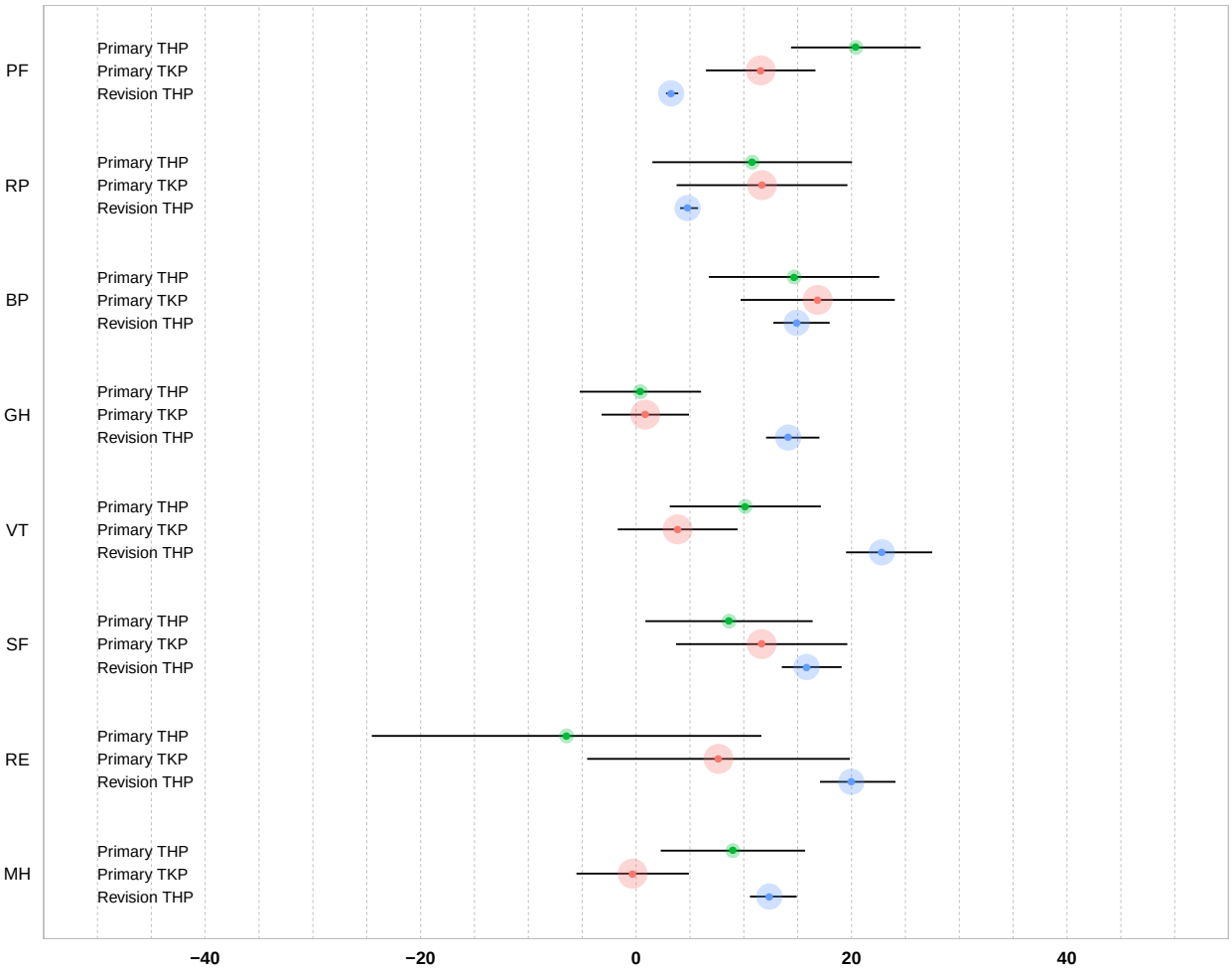


Figure 6.2: **Graph showing the minimal clinically important differences (MCIDs) in the domains of the Short-Form 36 (SF-36) at six months after primary total hip (THR)[169] and total knee replacement (TKR)[170] and revision THR.[199]** The size of the coloured circles represents the sample sizes used to estimate the MCID, and the error bars denote the 95% confidence intervals. PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

However, in order to visualise the precision of all MCID estimates, we calculated 95% confidence intervals of all MCID estimates, which were not presented in the original studies. These confidence intervals are presented in figure 6.2 and figure 6.3 on the next page.

The findings of this systematic review underline the need to identify MCIDs for each specific population. As can be seen from figure 6.2 and figure 6.3 on the next page, MCIDs differ both between SF-36 subscales and patient populations. The use of a

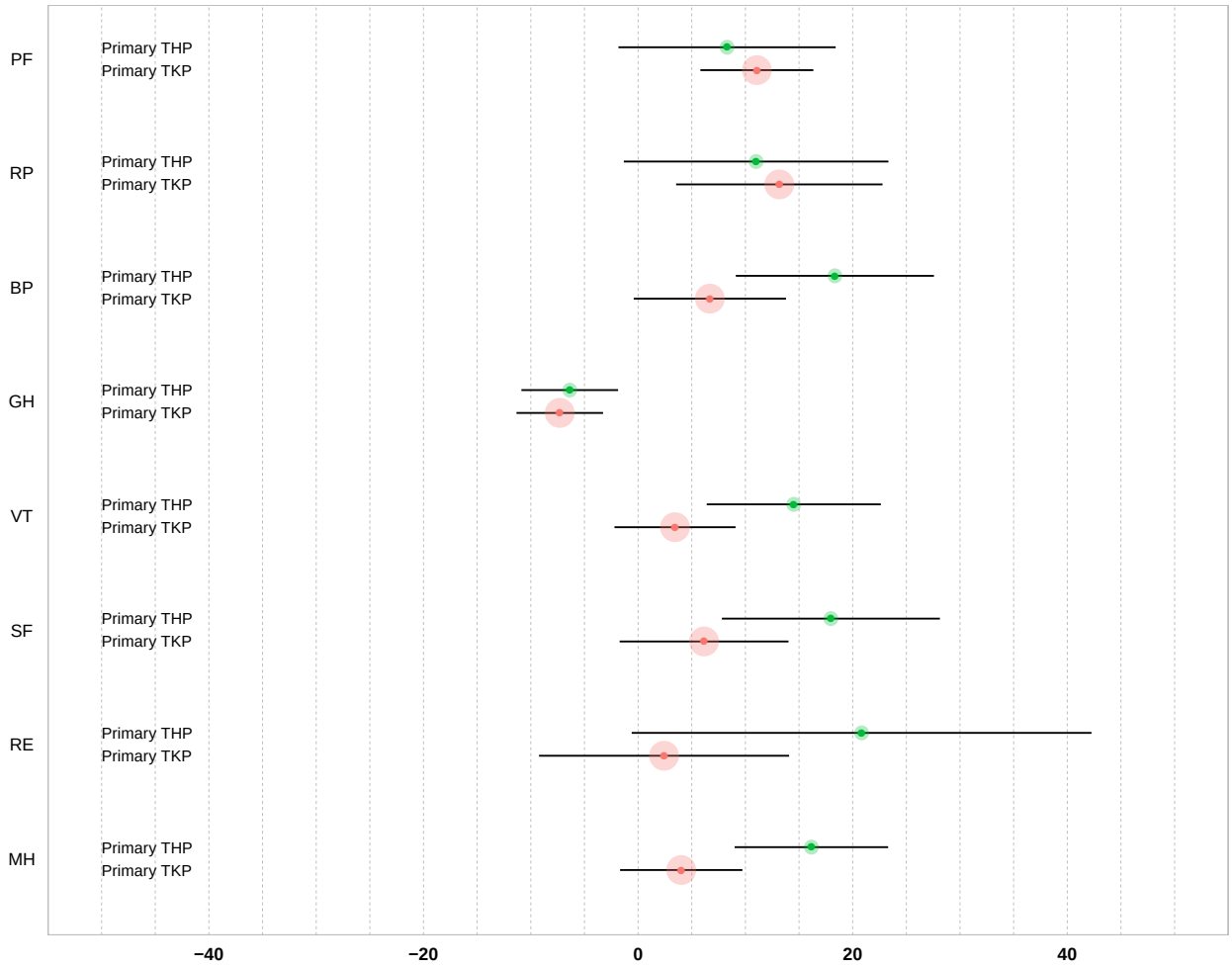


Figure 6.3: **Graph showing the minimal clinically important differences (MCIDs) in the domains of the Short-Form 36 (SF-36) at two years after primary total hip (THR)[169] and total knee replacement (TKR).[170]** The size of the coloured circles represents the sample sizes used to estimate the MCID, and the error bars denote the 95% confidence intervals. PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

"one-size-fits-all" MCID does not appear justified, as patients suffering from osteoarthritis of the hip and knee, which are regarded as similar disease entities, have different MCIDs in HRQoL.[200–202] In order to study patient-relevant improvements in HRQoL at the individual level in revision TKR patients, MCIDs need to be established in this particular population as well.

Limitations of the included studies include imprecision as a result of small sample sizes, the lack of validation of the MCID estimates and the rates of loss to follow-up. Anchor-based approaches in particular suffer from imprecision due to small sample

sizes, as this approach uses only a part of all data to estimate the MCID. A precise estimation of the MCID is further hampered by the clinical success of joint replacement: typically, one expects a large effect of THR or TKR.[203] The group sizes of patients who indicate that their condition has “somewhat improved” are therefore expected to be small, which contributes to an imprecise estimation of MCIDs. Unfortunately, there are only two ways to improve the precision of anchor-based MCID estimates: one can either perform larger studies, or pool study results in a meta-analysis. To date, the only studies that have established anchor-based MCIDs in HRQoL after primary THR or TKR were those of Quintana et al[169] and Escobar et al.[170] More research is required to improve the precision of MCIDs in HRQoL. Estimates with higher precision are generated by distribution-based approaches, which use data from the entire population to estimate the MCID. However, these approaches are criticised for the arbitrariness of the individual effect size standards.[192]

A strongly recommended method of determining MCIDs is by triangulation of multiple approaches.[193] None of the included studies has applied any form of additional validation, such as secondary anchor questions; all used a single approach. Besides a limited precision, caused by small group sizes, the accuracy of the MCID estimates might be limited as well due to this lack of additional validation.[204] Therefore, further research is needed to provide external validation of the established MCIDs in HRQoL.

However, until further research is performed, the MCID estimates of these three studies[169, 170, 199] are the best available estimates. Cautious use of these estimates should be encouraged in order to study improvement in HRQoL at the individual level, the most relevant outcome measure for individual patients encountered in clinical practice.

7

Willingness to Undergo Surgery Again Validated Clinically Important Differences in Health-Related Quality of Life after Total Hip and Knee Replacement Surgery

JC Keurentjes[1], M Fiocco[2], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

The Journal of Clinical Epidemiology. 2014; 67 (1), 114-120.

Abstract

Objective To determine Clinically Important Differences (CIDs) in Health-Related Quality of Life (HRQoL) after Total Hip (THR) or Total Knee Replacement (TKR), using the Short-Form 36 (SF36).

Methods SF36 scores were collected 2 weeks before and at 2-6 years after joint replacement in 586 THR and 400 TKR patients in a multi-center cohort study. We calculated distribution based CIDs (0.8 standard deviations of the preoperative score) for each SF36 sub-scale. Responders (patients with an improvement in HRQoL $\geq$ CID of a particular sub-scale) were compared to non-responders using an external validation question: willingness to undergo surgery again.

Results CIDs for THR/TKR were: Physical Functioning (PF): 17.9/16.7; Role-Physical (RP): 31.1/33.4; Bodily Pain (BP): 16.8/16.2; General Health (GH): 15.5/15.7; Vitality (VT): 17.3/16.7; Social Functioning (SF): 22.0/19.9; Role-Emotional (RE): 33.7/33.6; Mental Health (MH): 14.8/14.1. CIDs of PF, RP, BP and SF were validated by the validation question.

Conclusions Valid and precise CIDs are estimated of PF, RP, BP and SF, which are relevant HRQoL subscales for THR and TKR patients. CIDs of all other subscales should be used cautiously.

Introduction

Total Hip (THR) and Knee Replacement (TKR) alleviate pain and improve Health-Related Quality of Life (HRQoL) at the population level.[1] This information may not be meaningful for individual patients in clinical practice, who are interested in the likelihood of experiencing a meaningful improvement for the risk they take with an intervention.[191] Clinically Important Differences (CIDs), defined as a difference in scores of an outcome measure that is perceived by patients as beneficial or harmful,[167, 168] can be used to estimate the probability of achieving a meaningful improvement. Patients experience a meaningful improvement if their improvement $\geq$ the CID threshold; patients who improve less or deteriorate are considered non-responders.

As risks, costs and expected benefits vary widely between different interventions,[192] CIDs for a generic HRQoL instrument (e.g. the Short-Form 36;SF36) may vary across applications.[193] Minimal CIDs (MCIDs) after THR and TKR for the SF36 were recently summarized in a systematic review.[141, 169, 170] However, these estimates were not validated using external criteria.[141] Additionally, the relevance of a minimal improvement after THR or TKR is debatable, as one would generally expect a larger improvement after joint replacement.[203] Finally, the recommended anchor-based approach yielded imprecise CID estimates, which are not suitable for clinical practice. As large improvements in HRQoL are expected from joint replacement, the number of patients who rated their improvement after joint replacement as “somewhat better” was small, rendering imprecise CID estimates.

In order to overcome this limitation of anchor-based CID estimates in treatments with a large effect-sizes, such as joint replacements, we propose a new approach, combining efficient distribution-based CID estimation with anchor-based external validation. We used this approach to estimate CIDs in HRQoL after THR and TKR.

Methods

The current study is part of a multi-center follow-up study of HRQoL after THR or TKR (NTR2190). IRB approval was obtained from all participating centers, all patients gave written informed consent (CCMO-Nr:NL29018.058.09;MEC-Nr:P09.189). The data used in this report constitute a subset of patients, who underwent primary THR or TKR and have completed pre-operative and post-operative HRQoL questionnaires.

Assessments HRQoL was measured 2 weeks before TKR / THR and 1.5-6 years after surgery, using the Dutch SF36.[130, 133] The 36 items cover eight domains (physical function (PF), role physical (RP), bodily pain (BP), general health (GH), vitality VT), social function (SF), role emotional (RE), and mental health (MH)), for which a sub-scale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). Missing items were imputed according to Ware whenever possible.[197]

A validation question (VQ) was included in the questionnaire: “knowing what your hip or knee replacement surgery did for you, would you still have undergone this surgery (yes / no)?”. This validation question was previously used in a similar study, which validated WOMAC CIDs after THR and TKR.[203]

Outcome measures CIDs can be established using anchor-based or distribution-based methods.[192, 193, 195] In an achor-based approach, the target instrument is related to an independent measure (an anchor).[192] Typically, within-patient global change ratings (measured using a Likert-scale) are used as anchors; the CID is estimated by the mean improvement of patients who report that their condition is at least “somewhat better”.[195] In a distribution-based approach, the magnitude of the effect is related to a measure of variability of results.[192] Typically, Cohen’s effect-size benchmarks[196] are adapted for individual effect sizes, giving 0.3 or 0.5 times the standard deviation of the baseline score for a minimal CID and 0.8 times the standard deviation of the baseline score for a CID.[195]
In order to estimate CIDs, we chose the following, two-phased approach. In the first phase, we estimated the CID using a distribution-based approach. This approach generates a more precise estimate of the CID, because information from the entire cohort is used, instead of only a part of the population as is the case in anchor-based methods. In the second phase, the distribution-based CIDs were validated by the validation question VQ.

Statistical analyses Baseline characteristics were compared using descriptive statistics. Distribution-based CIDs in HRQoL of THR and TKR patients were calculated by multiplying the standard deviation of the untransformed sub-scale scores at baseline by 0.8, which indicates a large group change.[195] We validated the CIDs using the VQ. Each individual patients’ improvement (ie the postoperative score minus the preoperative score) was computed and compared to the CID. A 2 by 2 contingency table was constructed for each subscale of the VQ to display the numbers of individuals who had an improvement equal to or larger than the CID and gave positive or negative answers

to the VQ or had an improvement smaller than that of the CID and gave positive or negative answers to the VQ. For each contingency table, an odds ratio was calculated, which can be interpreted as the ratio of the odds of having experienced a CID when patients have expressed willingness to undergo surgery again, relative to the odds of not having experienced a CID when patients have expressed willingness to undergo surgery again. An odds ratio larger than 1 indicates that that particular CID is able to discriminate patients who answered the VQ positively from patients who answered the VQ negatively.

Sensitivity analyses: in order to check whether the odds ratios of the validation procedure were robust across different arbitrary CID threshold, we repeated all analyses using the following CID thresholds: 0.3 * SD and 0.5 * SD.

In order to verify whether the estimated CIDs are consistent across different subpopulations, we calculated the CIDs separately for different subgroups and compared these to the overall CID estimates. Subgroup CID estimates were calculated for strata of the following variables: sex, age (<65 vs $\geq$65 years old) and Charnley classification (Class A: patients in which the index operated hip or knee are affected only; Class B: patients in which the other hip or knee is affected as well; Class C: patients with a hip or knee replacement and other affected joints and/or a medical condition which affects the patients' ability to ambulate).[136, 137]

Results

Population Patient characteristics are presented in table 7.1 on the following page. 586 patients underwent THR and 400 underwent TKR. The average follow-up period was similar for THR and TKR patients (3.2 years (SD 1.1), both for THR and TKR). THR patients were slightly younger at joint replacement surgery (mean age at joint replacement (SD): THR 66 (10.6); TKR 69.1 (9.6)). The proportion of males was similar (THR 34.1%, TKR 33.3%). TKR patients had a higher mean BMI and were more often obese or morbidly obese. The majority underwent joint replacement for primary OA.

2206 patients underwent primary joint replacement and were eligible for inclusion in this

	Primary THR: n = 586	Primary TKR: n = 400
Mean Follow-up Years (SD);	3.2 (1.1);	3.2 (1.1);
Median (IQR)	3.0 (2.3 – 4)	2.9 (2.3 – 4)
Mean Age (SD);	66 (10.6);	69.1 (9.6);
Median (IQR)	67 (60.4 – 73.6)	70.4 (63.1 – 76.5)
$\leq$ 50 Years (%)	46 (7.8)	9 (2.3)
51 - 60 Years (%)	95 (16.2)	65 (16.5)
61 - 70 Years (%)	221 (37.7)	118 (30.0)
71 - 80 Years (%)	187 (31.9)	156 (39.7)
$>$ 80 Years (%)	37 (6.3)	45 (11.5)
Number of Men (%)	200 (34.1)	132 (33.3)
Mean BMI* (SD);	27.1 (4.2);	29.2 (4.9);
Median (IQR)	26.6 (24.2 – 29.4)	28.5 (25.8 – 32)
$<$25 (%)	191 (34.3)	68 (18.0)
25-30 (%)	243 (43.6)	168 (44.4)
30-35 (%)	98 (17.6)	97 (25.7)
$>$35 (%)	25 (4.5)	45 (11.9)
Indication for Joint Replacement:		
Osteoarthritis (%)	501 (86.2)	354 (89.4)
Rheumatoid arthritis (%)	13 (2.2)	26 (6.6)
Other (%)	68 (11.7)	16 (4.0)

Table 7.1: **Patient Characteristics.** *Measured at follow-up.

follow-up study. 285 patients did not complete all pre-operative questionnaires and 63 patients died, leaving 1858 patients with primary joint replacement eligible. 986 patients agreed to participate and returned the questionnaires sufficiently completed (response rate: 53%). Non-responding THR patients were on average 3.95 years older than participants (95%CI: 2.6-5.3 years); Non-responding TKR patients were on average 3.31 years older than participants (95%CI: 2.0-4.7 years). The proportion of males was similar in participants and non-responders.

Phase 1: CID Estimation The mean preoperative scores of the SF36 subscales are presented in table 7.2 on the next page and 7.3 on the facing page. For THR patients, the following improvement in HRQoL scores after joint replacement constitutes a CID: physical functioning: 17.9; role-physical 31.1; bodily pain: 16.8; general health: 15.5; vitality: 17.3; social functioning: 22.0; role-emotional: 33.7; mental health: 14.8.

SF36 Subscale	Mean Pre-Operative Score (SD)	Clinically Important Difference (95%CI)
Physical Functioning	40.1 (22.3)	17.9 (16.9–19.0)
Role-Physical	30.9 (38.9)	31.1 (29.4–33.1)
Bodily Pain	40.3 (20.9)	16.8 (15.8–17.8)
General Health	67.8 (19.3)	15.5 (14.6–16.4)
Vitality	61.0 (21.6)	17.3 (16.3–18.4)
Social Functioning	65.6 (27.5)	22.0 (20.8–23.4)
Role-Emotional	68.9 (42.2)	33.7 (31.8–35.9)
Mental Health	74.3 (18.5)	14.8 (14.0–15.7)

Table 7.2: **Pre-operative HRQoL and CIDs in HRQoL of Primary THR.**

SF36 Subscale	Mean Pre-Operative Score (SD)	Clinically Important Difference (95%CI)
Physical Functioning	40.3 (20.8)	16.7 (15.5–18.0)
Role-Physical	38.8 (41.8)	33.4 (31.2–36.0)
Bodily Pain	44.9 (20.3)	16.2 (15.1–17.5)
General Health	62.8 (19.7)	15.7 (14.7–16.9)
Vitality	60.9 (20.9)	16.7 (15.6–18.0)
Social Functioning	70.5 (24.9)	19.9 (18.6–21.5)
Role-Emotional	68.8 (42.0)	33.6 (31.3–36.2)
Mental Health	73.5 (17.7)	14.1 (13.2–15.2)

Table 7.3: **Pre-operative HRQoL and CIDs in HRQoL of Primary TKR.**

For TKR patients, the following improvement in HRQoL scores after joint replacement constitutes a CID: physical functioning: 16.7; role-physical 33.4; bodily pain: 16.2; general health: 15.7; vitality: 16.7; social functioning: 19.9; role-emotional: 33.6; mental health: 14.1.

Phase 2: Validation Box plots of the improvement in eight dimensions of HRQoL after joint replacement in relation to the CID threshold for each dimension, stratified by the response to the validation question, are shown in figure 7.1 on the next page for THR patients and in figure 7.2 (p. 101) for TKR patients. THR patients who reported having a larger improvement in physical functioning, role-physical, bodily pain, general heath, social functioning and role-emotional than the CIDs, had also expressed willingness to undergo surgery again more often. These findings are also reflected in the odds ratios, which are larger than 1 (table 7.4 (p. 102)). TKR patients who reported having a larger improvement in physical functioning, role-physical, bodily pain and social functioning than the CIDs, had also expressed willingness to undergo surgery again more often.

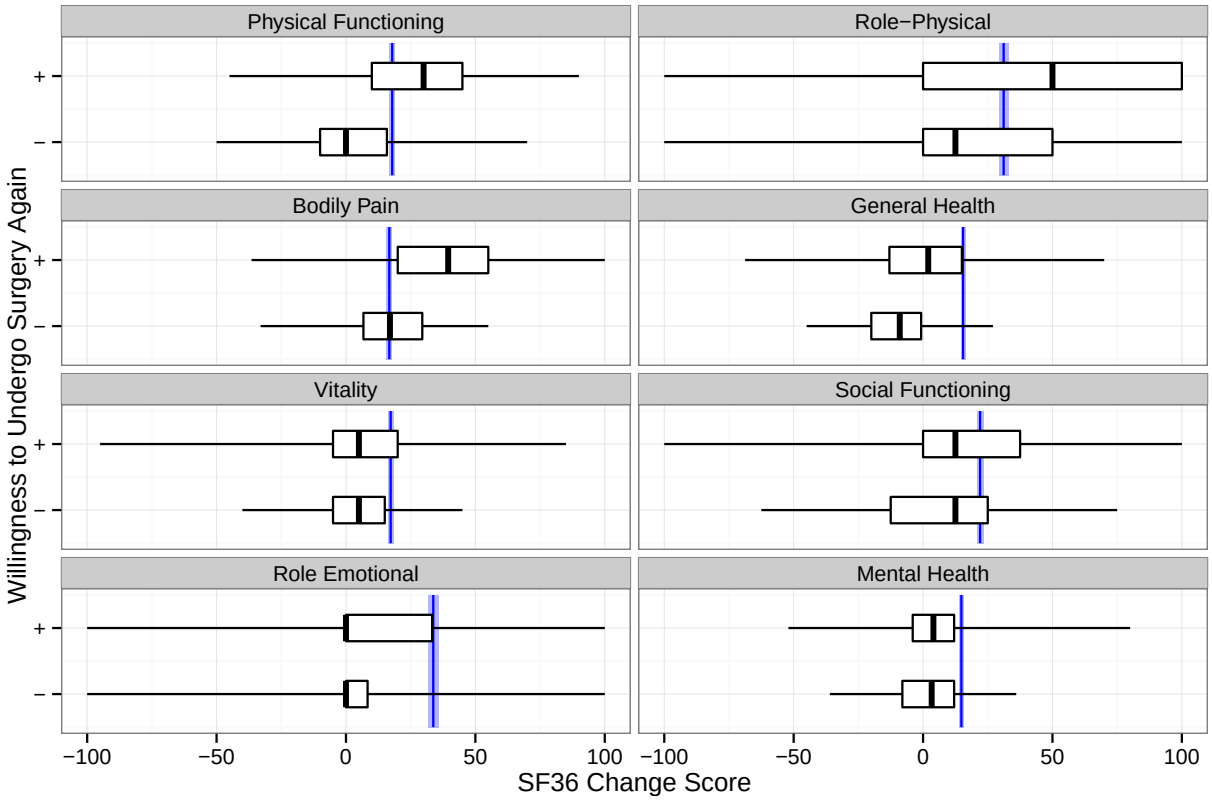


Figure 7.1: **Improvement in HRQoL after THR per Validation Question.** The vertical blue lines indicate the CID of each sub-scale with its confidence interval shown in purple; the boxplots indicate the median, IQR and range of patients, who answered the Validation Question positively and negatively.

These findings are also reflected in the odds ratios, which are larger than 1 (table 7.4 (p. 102)). All contingency tables from which these odds ratios were calculated, can be found online at *http://www.ncbi.nlm.nih.gov/pubmed/23850406*.

Sensitivity analyses showed similar odds ratios for different CID thresholds, indicating a robustness of the association between achieving a CID and expressing willingness to undergo surgery again, for different thresholds (contingency tables can be found online at *http://www.ncbi.nlm.nih.gov/pubmed/23850406*). CIDs were similar for men and women, for patients younger and older than 65 years and for different Charnley classes (data not shown).

Discussion

We have established CIDs in HRQoL after THR and TKR and have validated these estimates using a relevant validation question. The CID estimates of physical functioning,

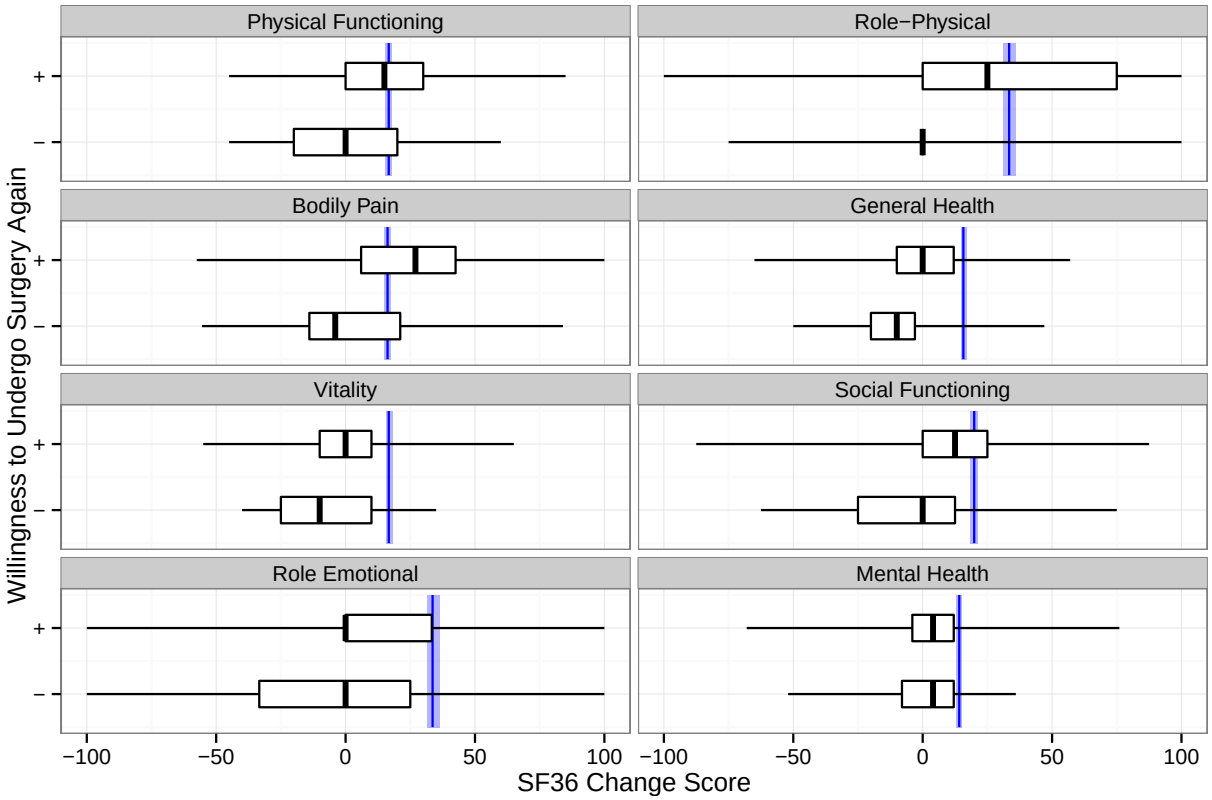


Figure 7.2: **Improvement in HRQoL after TKR per Validation Question.** The vertical blue lines indicate the CID of each sub-scale with its confidence interval shown in purple; the boxplots indicate the median, IQR and range of patients, who answered the Validation Question positively and negatively.

role-physical, bodily pain and social functioning are both precise (judged by the narrow confidence intervals)[204] and accurate (due to the validation procedure), enabling further research in HRQoL gains after THR or TKR at the individual level. CIDs of all other SF36 subscales should be used cautiously.

A limitation of our study is the variable length of follow-up, which ranges from 1.5 to 6 years after surgery. CIDs might be different for patients with different lengths of follow-up. However, recent evidence suggests that gains in HRQoL are sustained up to 7 years after joint replacement.[147, 148]

In establishing a CID for a specific outcome measure, it is recommended to use multiple approaches and triangulation of methods.[193] Anchor-based approaches are preferred, as these are explicitly attached to observed mean changes. Distribution-based approaches have been criticized for being non-intuitive and arbitrary in the choice

SF36 Subscale	Odds Ratio (95%-Confidence Interval) THR	TKR
Physical Functioning	5.86 (3.13–11.7)	1.80 (0.78–4.52)
Role Physical	2.08 (1.13–3.95)	2.98 (1.19–9.20)
Bodily Pain	3.30 (1.81–5.98)	4.72 (2.07–11.8)
General Health	4.92 (1.76–21.2)	1.26 (0.46–4.51)
Vitality	1.11 (0.59–2.22)	0.78 (0.32–2.20)
Social Functioning	1.89 (1.02–3.62)	3.35 (1.25–11.9)
Role Emotional	2.84 (1.11–9.83)	0.68 (0.29–1.81)
Mental Health	1.06 (0.55–2.18)	0.95 (0.39–2.70)

Table 7.4: **Odds ratios of attaining a CID and being willing to undergo surgery again for primary THR and TKR.** An odds ratio >1 indicates that that particular CID is able to discriminate patients who answered the VQ positively from patients who answered the VQ negatively.

of the individual effect size standards.[192, 195] However, anchor-based methods might not be feasible in THR or TKR. Quintana and Escobar advise against using their MCIDs due to the imprecision of these estimates.[169, 170] To augment the precision of these estimates, one would need very large cohorts. For instance, Quintana started with 586 eligible THR patients and ended with 33 patients at two years follow-up, who described their status as “somewhat better”. In order to end up with 100 patients and achieve a more precise CID, approximately 1750 eligible patients would be necessary. Additionally, arbitrary thresholds also play a role in anchor-based approaches. Chesworth et al have defined the CID as the mean improvement in the WOMAC score of patients who indicated +5 on a 15-point general transition Likert scale.[203] Similar to the arbitrary effect sizes of Cohen, +5 might be reasonable, but remains an arbitrary choice.

Our new approach overcomes these limitations in treatments with large effect-sizes. In order to ensure precise estimates, we estimated CIDs using the distribution-based approach. This approach uses data of the entire cohort, enhancing the precision of the estimate as compared to anchor-based approaches. To overcome the non-intuitivity of the distribution-based approach, we have validated the CID estimates using a patient relevant external criterion. Clinical meaningfulness is regained by means of the odds ratios.

Why are CIDs useful in treatments with large effect-sizes? Although on average patients improve markedly after THR or TKR, not all patients benefit from these surgeries. Persistent pain is reported in 9% of THR patients and 20% of TKR patients at long term follow-up.[11] Additionally, up to 30% of patients are dissatisfied with the surgical results.[12–15] Therapeutic options are limited in patients with persistent pain or dissatisfaction after joint replacement: the outcome of revision surgery performed without a specific mechanical or physiological indication is highly unpredictable. Furthermore, revision surgery is associated with a higher probability of orthopaedic and medical complications. Unfulfilled patient expectations are thought to play a crucial role in unfavourable outcomes after joint replacement.[166] CIDs might bridge the gap between patient expectation and satisfaction. Using CID thresholds, it will be possible to predict the probability of a relevant improvement in various relevant areas of HRQoL, using clinical prediction models. These predictions for individual patients could be made before surgery has taken place, and could form a solid base for expectation management. Such a tailored approach could lower the probability of unfavourable outcomes after joint replacement in future patients.

8

Patient Acceptable Symptom States After Total Hip or Knee Replacement at Mid-term Follow-up

JC Keurentjes[1], FR Van Tol[1], M Fiocco[2], C So-Osman[3], R Onstenk[4], AW Koopman-Van Gemert[5], RG Pöll[6], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Sanquin Blood Supply, South West Region.

4 Orthopaedic Surgery, Groene Hart Hospital, Gouda.

5 Anaesthesiology, Albert Schweitzer Hospital, Dordrecht.

6 Orthopaedic Surgery, Slotervaart Hospital, Amsterdam.

Bone & Joint Research. 2014; 3 (1), 7-13.

Abstract

Objectives To define Patient Acceptable Symptom State (PASS) thresholds for the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) at mid-term follow-up.

Methods In a prospective multicenter cohort study, OHS and OKS were collected at an average of 3 years follow-up, combined with a Numeric Rating Scale of Satisfaction (NRSS) and an external validation question (VQ): willingness to undergo surgery again.

Results 550 patients underwent THR and 367 underwent TKR. Receiver operating characteristic (ROC) curves identified a PASS threshold of 42 for the OHS after THR and 37 for the OKS after TKR. THR patients with an OHS$\geq$42 and TKR patients with an OKS$\geq$37 had a higher Numeric Rating Scale of Satisfaction and a larger odds of being willing to undergo surgery again.

Conclusions PASS thresholds appear larger at mid-term follow-up than at 6 months after surgery. Without external validation, we would advise against using these PASS thresholds as absolute thresholds in defining whether or not a patients has attained an acceptable symptom state after THR/TKR.

Introduction

Several distinct types of outcome measures are of interest in orthopaedic surgery. The time to a certain event, such as revision surgery, has historically been the principal outcome of interest in joint replacement patients.[116] In recent years, patient-reported outcome measures (PROMs) have become popular, allowing the assessment of the clinical outcome of joint replacement from the patient's perspective.[205] PROMs can be summarised in numerous ways. In the orthopaedic literature, mean scores of the study population are frequently presented. The mean pre-operative score provides information on the "average" patient before surgery. Similarly, the mean post-operative score provides information on the "average" patient after surgery, and the mean change in these scores provides information on the improvement (or deterioration) experienced by the "average" patient, who does improve substantially after joint replacement.[1] However, a large proportion of joint replacement patients suffer from persisting pain,

or are dissatisfied with the surgical results.[11, 15, 166] Data regarding the mean improvement after joint replacement mainly report the improvement of many patients with successful outcomes, but can neglect patients with suboptimal outcomes, making it of limited use for individual patients encountered in clinical practice.

Patient Acceptable Symptom States (PASS) and Minimal Clinically Important Differences (MCIDs) are two complementary constructs, which allow a more individualised approach to the analysis of PROMs.[193, 206, 207] PASS is defined as an outcome score threshold of the post-operative score, above which a patient is defined as experiencing a satisfactory outcome, and below which an unsatisfactory outcome is experienced. MCID is defined as the minimum amount of improvement between pre- and post-operative scores that a patient should experience after a specific intervention in order to have achieved a minimally important difference. PASSs and MCIDs allow estimation of the probability of a satisfactory outcome or a relevant improvement. These probabilities are relevant for individual patients, encountered in clinical practice, who either do or do not achieve an acceptable state or experience a relevant improvement.[191] Recently, PASSs have been estimated for the Oxford Hip Score (OHS)[208] and Oxford Knee Score (OKS)[209] at short-term follow-up.[210] An important issue is whether the chosen follow-up period of six months after joint replacement is adequate. A recent systematic review has suggested that patients may not have fully recovered at six months after THR.[149] Thresholds that define whether or not patients have achieved an acceptable symptom state, such as the PASS, may therefore differ between patients who are still recovering from their surgery and patients who have recovered fully. Therefore, we questioned whether PASS thresholds are different at mid-term compared with short-term follow-up. We questioned whether the OHS and OKS are correlated to patient satisfaction at mid-term follow-up. Additionally, we questioned whether responders (i.e., patients who have an acceptable symptom state according to the PASS) are more satisfied than non-responders. Finally, we questioned whether responders were more likely to be willing to undergo surgery again, compared with non-responders.

Materials and Methods

The current study is part of a multicentre cohort study of health-related quality of life (HRQoL) after THR/TKR (NTR2190), performed from August 2010 to August 2011.[21, 27, 28] Institutional Review Board approval was obtained from all participating centres and all patients gave written informed consent (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189). It concerned the clinical follow-up of a multi-centre randomised controlled clinical trial, comparing the use of the drug erythropoietin and two re-infusion techniques of autologous blood in order to decrease allogenic blood transfusions (Netherlands Trial Register: NTR303). In this trial, 2442 primary and revision hip or knee replacements in 2257 patients were included between 2004 and 2009. All patients who participated in the randomised controlled trial completed pre-operative HRQoL questionnaires, underwent primary THR or TKR and who were alive at the time of inclusion for the present follow-up study, were eligible for inclusion. The first joint replacement was selected for inclusion in the follow-up study for patients who participated more than once in the previous study. Records of the financial administration of all participating centres were checked in order to ascertain that all eligible patients were alive before being approached by the first author (JCK). For the present follow-up study, all eligible patients were first sent an invitation letter signed by their treating orthopaedic surgeon, an information brochure and a reply card. Patients who indicated that they were willing to participate were sent a questionnaire. Patients who did not respond to the first invitation within four weeks were sent another invitation letter. Those who did not respond to this second invitation were contacted by telephone by the first author. Patients who did not return their questionnaire within four weeks were also contacted by telephone by the first author. The data used in this report constitute a subset of patients who completed post-operative questionnaires.

Outcome measures: We measured the overall satisfaction with the outcome of surgery on a numeric rating scale (NRS), which ranged from 0 (extremely dissatisfied) to 10 (extremely satisfied). We added a validation question to the questionnaire, which took the following form: “Knowing what your hip or knee replacement surgery did for

you, would you still have undergone this surgery?", with dichotomous answers of 'yes' vs 'no'. This validation question was previously used in a similar validation study of clinically important differences after THR and TKR.[203]

Joint-specific PROMs were measured using the OHS for THR patients and the OKS for patients undergoing TKR, both of which were translated and validated in Dutch.[211, 212] Each questionnaire comprises 12 questions regarding pain and functioning of the hip or knee during the previous four weeks. Each question is answered on a five-point Likert scale, and an overall score is calculated by summarising the responses to each of the 12 questions. This sum score ranges from 0 to 48, where 0 indicates the most severe symptoms and 48 the least severe symptoms.

Potential confounders included age at joint replacement, gender, body mass index (BMI), indication for joint replacement (osteoarthritis (OA) vs other), patient-reported Charnley classification of comorbidity (A, patients in which the index operated hip or knee are affected only; B, patients in which the other hip or knee is affected as well; C, patients with a hip or knee replacement and other affected joints and/or a medical condition which affects the patients' ability to ambulate)[136, 137] and pre-operative HRQoL. HRQoL was measured pre-operatively using the Short-Form (SF-)36,[197] which is translated and validated in the Dutch language.[133] The 36 items cover eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health), for which a sub-scale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). These sub-scales can be summarised in a physical component summary (PCS) and a mental component summary (MCS).

Statistical Analysis: We performed descriptive analyses of the patients' baseline characteristics. All analyses were performed separately for THR and TKR, as MCIDs have been shown to differ considerably between these surgical interventions.[141] The correlation between OHS or OKS and NRS for satisfaction was calculated using Pearson's correlation coefficient. Receiver operating characteristic (ROC) curves were used to identify thresholds for OHS/OKS scores at mid-term follow-up, which are

associated with an acceptable level of patient satisfaction with joint replacement. An acceptable level of patient satisfaction was defined as a NRS for satisfaction $\geq$ 5, which is the equivalent of a visual analogue scale satisfaction score $\geq$ 50.[18, 205] This particular threshold has been used previously to compare satisfied and dissatisfied patients after joint replacement.[210] The chosen PASS thresholds were equivalent to the point at which sensitivity and specificity were closest.[213] The 95% confidence intervals (CI) around the thresholds were estimated using percentile bootstrap methods, based on 1000 random samples with replacement from the original data. In order to explore whether the found thresholds are consistent across subgroups, we identified separate thresholds for subgroups based on the following variables: length of follow-up (< 3 years vs $\geq$ 3 years), gender, age (< 70 years vs $\geq$ 70 years), BMI (< 30 kg/m2 vs $\geq$ 30 kg/m2), Charnley classification (A/B vs C), SF-36 PCS (< 50 vs $\geq$ 50) and SF-36 MCS (< 50 vs $\geq$ 50).

Based on the overall PASS thresholds, we divided patients into responders (those with an OHS or OKS $\geq$ the PASS threshold) and non-responders (OHS/OKS < PASS threshold). We compared the mean NRS for satisfaction between responders and non-responders separately for THR and TKR patients, using three different models. In the first model, we calculated the mean NRS for satisfaction of all responders and the mean NRS of all non-responders, stratified by centre. In the second model, we performed linear mixed model regression analyses, with age and gender as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length. The final model consisted of linear mixed model regression analyses, with age, gender, BMI, Charnley classification, indication (OA vs other), and pre-operative SF-36 PCS and MCS as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length.

Finally, we compared the odds of responding the validation question positively between responders and non-responders, using three different models. In the first model, we calculated crude odds ratios. In the second model, we performed logistic mixed model regression analyses, with age and gender as fixed effects and the centre as a random

effect, while stratifying for quartile of follow-up length. In the final model, we performed logistic mixed model regression analyses, with age, gender, BMI, Charnley classification, indication (OA vs other), and pre-operative SF-36 PCS and MCS as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length.
All analyses were performed using R v2.14.1.[43]

Results

A total of 550 patients underwent THR and 367 underwent TKR (see study flowchart in figure 1.1 (p. 5)). Patient characteristics are described in Table 8.1 on the following page. The mean follow-up was similar for THR and TKR patients, at 3.2 years (1.5 to 6.0) and 3.2 years (1.3 to 6.0), respectively. THR patients were slightly younger at joint replacement surgery than TKR patients. The proportion of males was similar. TKR were more often obese or morbidly obese. The majority of THR and TKR patients underwent joint replacement for primary OA.

The mean and median OHS scores at mid-term follow-up were 41.5 (sd 7.93) and 44 (interquartile range (IQR) 39 to 47), respectively. The mean NRS for satisfaction was 8.55 (sd 2.19) and 94.7% (521 of 550) of all THR patients were satisfied (defined as NRS $\geq$5). The mean and median OKS scores at mid-term follow-up were 39.1 (sd 9.04) and 42 (IQR 35 to 46), respectively. The mean NRS for satisfaction was 8.07 (sd 2.61) and 90.7% (333 of 367) of all TKR patients were satisfied.
The NRS correlated with both the OHS (Pearson correlation coefficient 0.52 (95% CI 0.46 to 0.58)) and the OKS (Pearson correlation coefficient 0.64 (95% CI 0.57 to 0.69)), indicating a strong correlation.

ROC curves of OHS thresholds and OKS thresholds are shown in Figures 8.1 (p. 113) and 8.2 (p. 114). The OHS ROC curve revealed a PASS threshold of 42, with a sensitivity of 67.0% and a specificity of 65.5%. The area under the ROC curve (AUC) was 0.72 (95% CI 0.60 to 0.84). The OKS ROC curve revealed a PASS threshold of 37, with a sensitivity of 76.3% and a specificity of 76.5%. The AUC was 0.83 (95% CI 0.74 to 0.93). ROC curves of subgroups showed variation in the thresholds found (Tables II and

	Primary THR: n = 550	Primary TKR: n = 367
Mean Follow-up Years (SD); Range	3.2 (1.1); 1.5 – 6.0	3.2 (1.1); 1.3 – 6.0
Mean Age at Joint Replacement (SD)	65.9 (10.5)	68.7 (9.64)
% Men:	34,2	33,3
BMI*		
% <25:	34,3	17,9
% 25-30:	42,6	44,5
% 30-35:	17,7	23,4
% >35:	5,82	14,2
Indication for Joint Replacement		
% Osteoarthritis:	86,3	89
Patient-reported Charnley Classification*		
% A:	23,1	13,9
% B:	14,3	10,2
% C:	62,6	75,9
Median Preoperative SF36 Summary Scores (IQR)		
Physical Component Summary	39.8 (34.1 – 45.3)	41.3 (35.0 – 47.3)
Mental Component Summary	54.8 (45.6 – 60.0)	54.1 (45.4 – 59.1)

Table 8.1: **Patient Characteristics.** *Measured at follow-up.

III). The variation appears larger in OHS thresholds than in OKS thresholds.

The mean NRS for satisfaction was significantly higher in responders than in non-responders, both for THR and TKR (Table 8.4 (p. 116)). Both models showed a mean difference between responders and non-responders of approximately two points for THR patients and three points for TKR patients. Responders were more likely to be willing to undergo surgery again (Table 8.5 (p. 117)). All models showed odds ratios of approximately 7, indicating a seven-times higher odds of willingness to undergo surgery again in responders versus non-responders, while controlling for confounding (Table 8.5 (p. 117)).

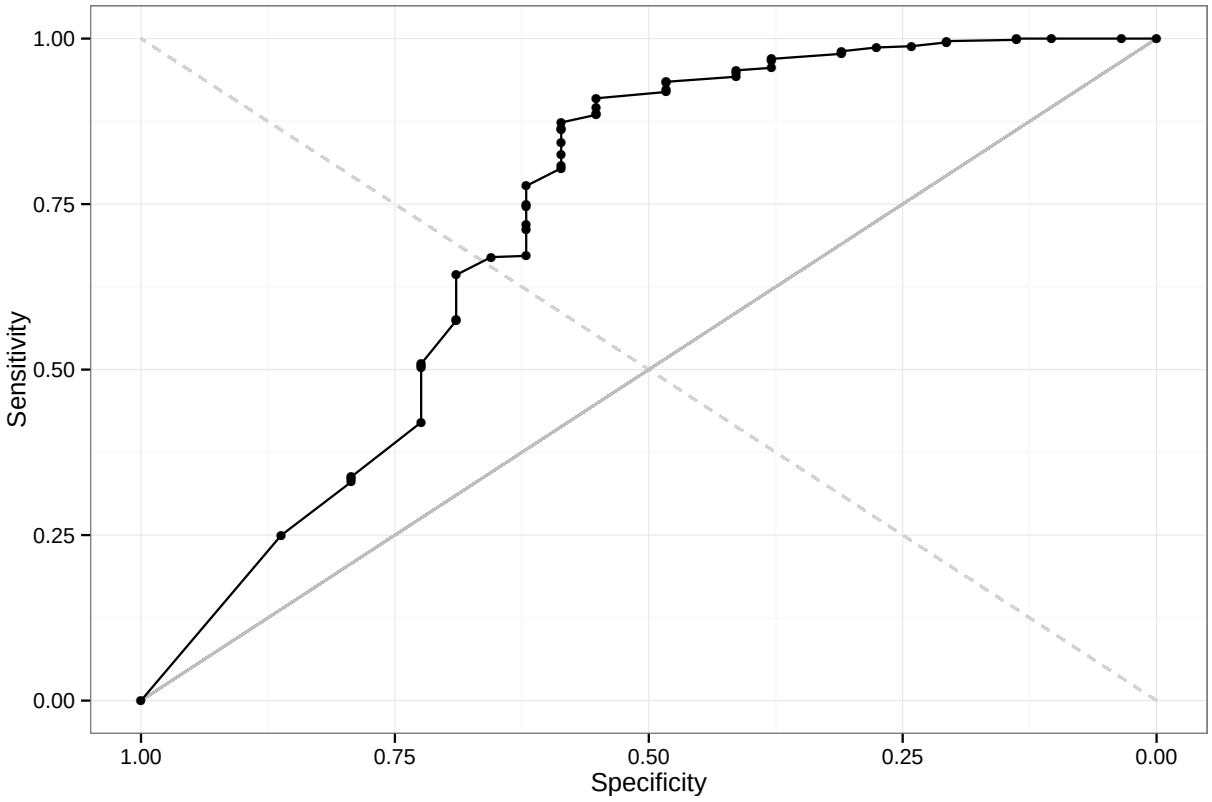


Figure 8.1: **Receiver Operating Characteristic (ROC) curves to identify thresholds for the mid-term follow-up Oxford Hip Score (OHS) associated with mid-term satisfaction with surgery.** Area under ROC curve (AUC): 0.72 (95%-CI: 0.60–0.84). Sensitivity: 67.0%, Specificity: 65.5%.

Discussion

PASS thresholds for the OHS and OKS are considerably higher at mid-term follow-up than those at six months post-operatively. The multiple approaches in validating the PASS thresholds and the rigorous efforts to minimise confounding are the main strengths of this study. All approaches show that the thresholds of 42 points for the OHS and 37 points for the OKS can discriminate between successful and less successful patient outcomes after THR or TKR in this study population, according to the overall satisfaction assessment and the willingness to undergo surgery again.

A limitation of our study is that we did not measure the OHS or OKS pre-operatively. Consequently, we could not investigate whether the PASS thresholds are valid across strata of baseline OHS or OKS scores. As a surrogate measurement of pre-operative joint functioning, we investigated differences in PASS thresholds in strata of the pre-

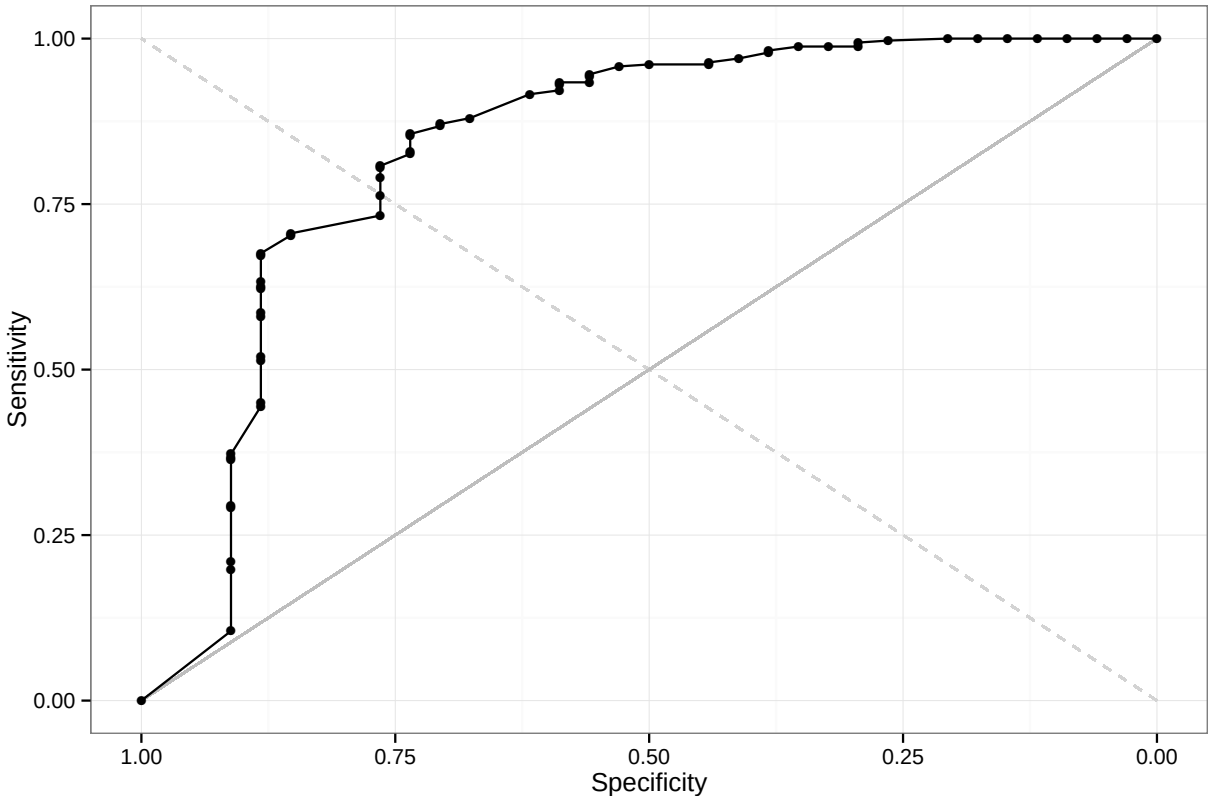


Figure 8.2: **Receiver Operating Characteristic (ROC) curves to identify thresholds for the mid-term follow-up Oxford Knee Score (OKS) associated with mid-term satisfaction with surgery.** Area under ROC curve (AUC): 0.83 (95%-CI: 0.74–0.93). Sensitivity: 76.3%, Specificity: 76.5%.

operative physical and mental component summaries of the SF-36, which only had a small effect. Furthermore, other evidence from the same research group suggests that baseline OHS or OKS values are poor predictors of overall patient satisfaction with the outcome of the joint replacement.[210, 214]

Another limitation of our study is the broad range in follow-up length. In order to account for this range, we stratified our analysis per quartile of follow-up period. Although a residual effect of follow-up length cannot be excluded, we do not think this is very plausible, as recent evidence suggests that after full recovery has taken place, the improvement in joint function is sustained throughout mid-term follow-up.[148]

Demographically, our study population is similar to that of Judge et al.[210] Cultural differences cannot be excluded from explaining the differences found in PASS thresholds, although this is unlikely, given the resemblance of English and Dutch urban joint

	Thresholds (95% CI)	Satisfactory Symptom State: n (%)
Entire Population	42 (38.5 – 44.2)	359 (65.3)
Follow-up $<$ 3 yrs	45 (39.6 – 46.5)	134 (47.2)
Follow-up $\geq$ 3 yrs	39 (30.5 – 42.5)	184 (69.2)
Males	47 (44.2 – 47.5)	85 (45.2)
Females	38 (30.5 – 41.2)	258 (71.3)
Age $<$ 70 yrs	44 (39.1 – 45.5)	204 (60.0)
Age $\geq$ 70 yrs	40 (30.5 – 42.5)	127 (60.5)
BMI $<$ 30	43 (39.6 – 45.5)	280 (68.6)
BMI $\geq$ 30	36 (25.5 – 43.5)	55 (45.1)
Charnley Class A / B	45 (26.0 – 47.5)	150 (77.7)
Charnley Class C	41 (37.0 – 43.5)	177 (54.8)
PCS $<$ 50	43 (37.5 – 44.7)	287 (64.8)
PCS $\geq$ 50	47 (34.0 – 48.0)	94 (90.0)
MCS $<$ 50	38 (28.7 – 41.5)	94 (54.0)
MCS $\geq$ 50	45 (40.7 – 46.5)	156 (48.3)

Table 8.2: **Patient acceptable symptom state (PASS) score thresholds (95% Confidence Intervals) for Mid-term follow-up Oxford Hip Score and percentage of patients classified as reaching a satisfactory symptom state, per relevant subgroups.**

replacement patients. A more plausible explanation could be the difference in physical recovery. Patients who are fully recovered at mid-term follow-up could be less prone to be satisfied with lower OHS or OKS scores, as the probability of further improvement in physical functioning is small. Six months after joint replacement, patients might be more readily satisfied with suboptimal OHS or OKS scores, as the speed of recovery is quite high.

Conceptually, MCIDs and PASSs are complementary. Both approach an individual patient's health state, but from a slightly different angle. In MCIDs, the emphasis is on whether or not an individual has improved after a certain therapy.[215] In PASSs, the emphasis is on whether or not the achieved outcome is acceptable from the patients perspective.[215] Both MCIDs and PASS have gained in interest recently.[141, 216–218] PASS might be more important than the MCID, as Dougados[219] phrased eloquently: "It's good to feel better but it's better to feel good." Similar methodological difficulties are encountered both in MCIDs and PASSs: both approaches lead to a loss of power

	Thresholds (95% CI)	Satisfactory Symptom State: n (%)
Entire Population	37 (31.8 – 38.5)	271 (73.8)
Follow-up $<$ 3 yrs	34 (29.7 – 37.5)	144 (78.7)
Follow-up $\geq$ 3 yrs	38 (32.5 – 41.5)	131 (74.0)
Males	39 (33.5 – 44.4)	89 (73.6)
Females	33 (30.5 – 37.5)	171 (70.7)
Age $<$ 70 yrs	34 (29.7 – 38.1)	139 (77.7)
Age $\geq$ 70 yrs	38 (31.3 – 41.7)	140 (77.3)
BMI $<$ 30	38 (32.0 – 39.5)	170 (78.7)
BMI $\geq$ 30	35 (29.7 – 40.5)	84 (64.6)
Charnley Class A / B	39 (30.8 – 45.5)	53 (62.4)
Charnley Class C	36 (30.5 – 38.5)	180 (67.4)
PCS $<$ 50	35 (31.3 – 38.5)	172 (63.2)
PCS $\geq$ 50	39 (25.0 – 47.0)	23 (39.0)
MCS $<$ 50	33 (21.5 – 38.5)	108 (86.4)
MCS $\geq$ 50	38 (33.5 – 42.3)	147 (71.4)

Table 8.3: **Patient acceptable symptom state (PASS) score thresholds (95% Confidence Intervals) for Mid-term follow-up Oxford Knee Score and percentage of patients classified as reaching a satisfactory symptom state, per relevant subgroups.**

	Responders:	Non-responders:	Mean Difference Adjusted for Age and Sex*	Mean Difference Adjusted for All Potential Confounders**
THR	9.28 (9.26 – 9.29)	7.26 (7.22 – 7.31)	1.89 (1.53 – 2.25)	1.87 (1.47 – 2.26)
TKR	9.04 (9.01 – 9.06)	6.78 (6.71 – 6.84)	2.89 (2.40 – 3.38)	2.96 (2.44 – 3.48)

Table 8.4: **Comparison of Mean Numerical Rating Scale of Satisfaction between responders and non-responders according to the Patient acceptable symptom state (PASS) score thresholds for THR and TKR.** * Stratified by quartiles of follow-up length. ** Adjusted for Age, Sex, BMI, Charnley Classification, OA vs other indications for joint replacement, preoperative Physical Component Summary Scale and preoperative Mental Component Summary Scale and stratified by quartiles of follow-up length.

compared with the population-level mean difference, both approaches depend on population and contextual characteristics and there is no clear consensus on the optimal statistical approach.[167, 193, 206] Despite these difficulties, MCIDs and PASSs are the best tools available to analyse PROM data at the individual level.

In this study, we estimated PASS in OHS/OKS at mid-term follow-up in a Dutch population. We found evidence suggesting that PASSs are time-dependent. Besides

	Crude Odds Ratio	Odds Ratio Adjusted for Age and Sex*	Odds Ratio Adjusted for All Potential Confounders**
THR	7.64 (4.09 – 15.3)	6.97 (3.51 – 13.8)	8.53 (3.80 – 19.1)
TKR	7.28 (3.85 – 14.3)	7.92 (3.79 – 16.6)	7.73 (2.84 – 21.0)

Table 8.5: **Comparison of Willingness to Undergo Surgery Again between responders and non-responders according to the Patient acceptable symptom state (PASS) score thresholds for THR and TKR.** An Odds Ratio >1 indicates a larger odds of Willingness to Undergo Surgery Again in Responders than in Non-responders. * Stratified by quartiles of follow-up length. ** Adjusted for Age, Sex, BMI, Charnley Classification, OA vs other indications for joint replacement, preoperative Physical Component Summary Scale and preoperative Mental Component Summary Scale and stratified by quartiles of follow-up length.

being time-dependent, PASS might also be population-dependent, as different sub-groups had different PASS thresholds. Without any form of external validation at a similar follow-up period, we would advise against using these PASS thresholds as absolute thresholds in defining whether or not a patient has attained an acceptable symptom state after THR/TKR.

9

Hip and Knee Replacement Patients Preferred Pen-and-Paper Questionnaires

JC Keurentjes[1], M Fiocco[2], C So-Osman[3], R Onstenk[4], AW Koopman-Van Gemert[5], RG Pöll[6], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Sanquin Blood Supply, South West Region.

4 Orthopaedic Surgery, Groene Hart Hospital, Gouda.

5 Anaesthesiology, Albert Schweitzer Hospital, Dordrecht.

6 Orthopaedic Surgery, Slotervaart Hospital, Amsterdam.

Bone & Joint Research. 2 (11), 238-244.

Abstract

Introduction Electronic forms of data collection have gained interest in recent years. In orthopaedics, little is known about patient preference regarding pen-and-paper or electronic questionnaires. We aimed to determine whether patients undergoing total hip (THR) or total knee replacement (TKR) prefer pen-and-paper or electronic questionnaires and to identify variables that predict preference for electronic questionnaires.

Methods We asked patients who participated in a multi-centre cohort study investigating improvement in health-related quality of life (HRQoL) after THR and TKR using pen-and-paper questionnaires, which mode of questionnaire they preferred. Patient age, gender, highest completed level of schooling, body mass index (BMI), comorbidities, indication for joint replacement and pre-operative HRQoL were compared between the groups preferring different modes of questionnaire. We then performed logistic regression analyses to investigate which variables independently predicted preference of electronic questionnaires.

Results A total of 565 THR patients and 387 TKR patients completed the preference question. Of the THR patients, 81.8% (95% confidence interval (CI) 78.4 to 84.7) preferred pen-and-paper questionnaires to electronic questionnaires, as did 86.8% (95% CI 83.1 to 89.8) of TKR patients. Younger age, male gender, higher completed level of schooling and higher BMI independently predicted preference of electronic questionnaires in THR patients. Younger age and higher completed level of schooling independently predicted preference of electronic questionnaires in TKR patients.

Discussion The majority of THR and TKR patients prefer pen-and-paper questionnaires. Patients preferring electronic questionnaires differed from patients who preferred pen-and-paper questionnaires. Restricting the mode of patient-reported outcome measures to electronic questionnaires might introduce selection bias.

Introduction

Traditionally, the assessment of outcome in orthopaedics has focussed on technical aspects. In total hip (THR) or knee replacement (TKR), the cumulative incidence of revision surgery is often used to compare the outcome of different implants or surgical techniques.[116] The underlying assumption of the traditional orthopaedic approach is that the technical aspects are the most important determinants of clinical success. However, a technically well-performed joint replacement does not guarantee clinical success, as no information is provided on functional status and pain. Additionally, the indication for revision surgery varies widely between orthopaedic surgeons.[115] Patient-reported outcome measures (PROMs), defined as questionnaires that are completed by patients, provide complementary information as they give an impression of a patient's experience of the surgical procedure and their concerns with regard to health status, health-related quality of life (HRQoL) and the results of the treatment received.[220]

PROMs can be measured using traditional pen-and-paper questionnaires or various electronic counterparts, including touch screens,[221] personal digital assistants,[24, 222] tablets or mobile phones.[223] Expected advantages of electronic questionnaires include more complete data capturing, immediate availability of results and lower costs of administrating and entering data.[23, 24]
On the other hand, electronic questionnaires may induce selection bias. A meta-analysis performed in 2008 showed that mail surveys had higher response rates than those based online.[224] A recent randomised controlled trial, in which 2400 patients were randomised to receive either a pen-and-paper questionnaire or an internet-based questionnaire at four years after THR, revealed an enormous difference in response rate: 92% for the pen-and-paper group versus 49% for the internet-based group.[225] Selection bias can occur if the association between exposure and outcome differs between participants and all eligible patients.[23]

To our knowledge, no study has investigated patient preference for electronic questionnaires after THR and TKR. The majority of members of a senior citizens

club prefers electronic to pen-and-paper questionnaires.[226] Given the similar age of THR/TKR patients, we would expect a preference for electronic questionnaires. We aimed to estimate the proportion of patients who prefer pen-and-paper questionnaires to electronic questionnaires and to estimate predictors of electronic questionnaire preference.

Materials and Methods

The current study is part of a multi-centre cohort study of HRQoL after THR/TKR (NTR2190), performed from August 2010 to August 2011.[21, 27, 28, 227] Institutional review board approval was obtained from all participating centers, all patients gave written informed consent (CCMO-Nr:NL29018.058.09;MEC-Nr:P09.189). The data used in this report constitutes a subset of patients who underwent primary THR or TKR and who completed pre-operative HRQoL questionnaires and a question regarding their preference for a mode of questionnaire at a mean of three years (1.5 to 6) after surgery.

We performed this study in order to investigate the preference for a mode of questionnaire for future studies in HRQoL after THR or TKR in a Dutch population. A prerequisite for such future studies is that patients can participate without outpatient department visits, thereby facilitating participation and forestalling the occurrence of selection bias. We selected a web-based questionnaire as the most feasible electronic option. At follow-up, we asked all THR and TKR patients which mode of questionnaire they preferred: pen-and-paper questionnaires or web-based electronic questionnaires, each completed at home.

In order to judge whether patients who preferred pen-and-paper questionnaires differed from patients who preferred electronic questionnaires, we compared age, gender, highest completed level of schooling, body mass index (BMI) categories ($<$ 25 kg/m^2, 25 to 30 kg/m^2, 30 to 35 kg/m^2, $>$ 35 kg/m^2), comorbidity, indication for joint replacement (osteoarthritis vs other indications) and pre-operative HRQoL between both groups.
We have aggregated the levels of schooling into an approximation of the social classes,

on the assumption that level of schooling indexes the type of qualifications obtained, which in turn indicates the type of occupations available to the subject and hence their own adult social class.[21] Thus: University, Higher vocational education and Preparatory higher vocational and scientific education have been aggregated as indicating the professional and managerial social classes; Middle vocational education and Preparatory middle vocational education have been aggregated as indicating the skilled non-manual and manual social classes; and Lower vocational education, Elementary schooling and No formal education have been aggregated as indicating the semi- and unskilled manual social classes.

Comorbidity was measured using a patient-reported Charnley classification (A, patients in which the index operated hip or knee are affected only; B, patients in which the other hip or knee is affected as well; and C, patients with a hip or knee replacement and other affected joints and/or a medical condition which affects the patients' ability to ambulate).[136, 137]

HRQoL was measured two weeks before TKR/THR, using the Dutch version of the Short-Form 36 (SF-36).[133, 197] This questionnaire comprises 36 items covering eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional and mental health), for each of which a subscale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). Additionally, these scales are incorporated into two summary measures: a physical component summary (PCS) and mental component summary (MCS). Missing items were imputed whenever possible according to Ware.[197] We compared pre-operative PCS and MCS between both preference groups.

Statistical analyses: We performed all analyses separately for THR and TKR patients, as clinically important differences differ considerably between these patient groups. [141] We performed descriptive analyses of baseline patient characteristics. In order to predict which factors increased the probability of preference for electronic questionnaires, we performed multivariate mixed model logistic regression analyses. We considered the

following potential predictors: age, gender, highest completed level of schooling category, BMI category, Charnley classification of comorbidity, indication for joint replacement and pre- operative PCS and MCS scores. In the mixed model regression analyses, patient preference was the dependent variable, all potential predictors were included as fixed effects and center was included as a random effect. The explained variation was estimated using Nagelkerke's generalised r^2 and the discriminative ability was estimated using the area under the receiver operating characteristic (ROC) curve (AUC).[228] The extent of optimism in the r^2 and AUC estimates was estimated using bootstrap resampling (n = 1000 bootstrap samples). [229–231]
All analyses were performed using R, version 2.15.2.[43]

Results

Patient characteristics are shown in table 9.1 on the facing page. A total of 565 THR patients and 387 TKR patients completed the preference question. Pen-and-paper questionnaires were preferred by 462 THR patients (81.8% (95% confidence interval (CI) 78.4 to 84.7) and by 336 TKR patients (86.8% (95% CI 83.1 to 89.8)) (table 9.2 on the next page).

Patient characteristics per preference group are show in table 9.3 (p. 126) for THR patients and table 9.4 (p. 127) for TKR patients, respectively. THR patients preferring electronic questionnaires tended to be younger, more often male, more often obese, less comorbid, more often highly educated and had worse pre-operative physical health. Age, gender and highest completed level of education remained associated with mode of questionnaire preference while adjusting for age and gender (table 9.3 (p. 126)). TKR patients who preferred electronic questionnaires were younger, more often male, less often morbidly obese, less often Charnley class B and more often Charnley class C, more often highly educated and had worse pre-operative physical health. Age and highest completed level of education remained associated with mode of questionnaire preference while adjusting for age and gender (table 9.4 (p. 127)).

	Primary THR: n = 565	Primary TKR: n = 387
Mean Follow-up Years (SD); Range	3.20 (1.13); 1.5 – 6.0	3.14 (1.12); 1.3 – 6.0
Mean Age at Joint Replacement (SD)	65.9 (10.6)	68.9 (9.66)
% Men:	35.0	32.6
BMI*		
% <25:	34.3	17.8
% 25-30:	42.9	44.1
% 30-35:	17.1	23.8
% >35:	5.71	14.2
Indication for Joint Replacement		
% Osteoarthritis:	86.1	89.3
Patient-reported Charnley Classification*		
% A:	23.3	14.6
% B:	14.2	10.5
% C:	62.6	74.9
Highest Completed Level of Schooling:		
% University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education:	22.6	15.4
% Middle Vocational Education and Preparatory Middle Vocational Education:	36.6	35.5
% Lower Vocational Education, Elementary Schooling and No Formal Education:	40.7	49.1
Mean Preoperative SF36 Summary Scores (SD)		
Physical Component Summary	38.9 (9.61)	40.6 (9.53)
Mental Component Summary	51.8 (10.8)	51.5 (10.2)

Table 9.1: **Patient Characteristics.** *Measured at follow-up.

Joint Replacement:	Proportion (%, 95% CI):
Total Hip Replacement	462 / 565 (81.8%; 78.4 – 84.7)
Total Knee Replacement	336 / 387 (86.8%; 83.1 – 89.8)

Table 9.2: **Proportion Of Patients Who Prefer Pen And Paper Questionnaires To Electronic Questionnaires, Per Joint Replacement.**

Multivariate prediction of electronic questionnaire preference showed that lower age, male gender, higher completed level of schooling and higher BMI independently predicted preference of electronic questionnaires in THR patients (table 9.5 (p. 128)). In TKR patients, multivariate prediction of electronic questionnaire preference showed that lower age and higher completed level of schooling independently predicted preference

	Pen and Paper	Electronic	Age and Gender-adjusted Odds Ratio
Mean Follow-up Years (SD)	3.17 (1.13)	3.31 (1.11)	-
Mean Age at Joint Replacement (SD)	67.5 (9.45)	58.5 (12.2)	0.93 (0.91 – 0.95)
% Men:	30,3	56,3	0.35 (0.22 – 0.56)
BMI			
% $<$25:	35,4	29,3	ref.
% 25-30:	42,1	46,5	1.32 (0.75 – 2.32)
% 30-35:	17,6	15,2	0.88 (0.42 – 1.83)
% $>$35:	4,95	9,09	2.18 (0.84 – 5.69)
Indication for Joint Replacement			
% Osteoarthritis:	88,5	75,5	0.75 (0.40 – 1.43)
Patient-reported Charnley Classification			
% A:	23,2	23,5	ref.
% B:	13,2	18,4	1.28 (0.59 – 2.79)
% C:	63,6	58,2	1.10 (0.61 – 1.98)
Highest Completed Level of Schooling			
% University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education:	19,1	37	ref.
% Middle Vocational Education and Preparatory Middle Vocational Education:	34,3	46	0.82 (0.47 – 1.45)
% Lower Vocational Education, Elementary Schooling and No Formal Education:	46,6	17	0.24 (0.12 – 0.47)
Mean Preoperative SF36 Summary Scores (SD)			
Physical Component Summary	39.5 (8.87)	35.8 (11.9)	1.00 (0.97 – 1.03)
Mental Component Summary	51.2 (10.6)	54.8 (10.9)	1.01 (0.99 – 1.04)

Table 9.3: **Comparison of Patients Who Prefer Pen And Paper Questionnaires With Patients Who Prefer Electronic Questionnaires: Total Hip Replacement.**

of electronic questionnaires (table 9.6 (p. 129)). The prediction model of Electronic questionnaire preference in THR patients had an r^2 of 0.31 with an optimism estimate of 0.04, yielding an optimism-corrected r^2 estimate of 0.27. The Area under the ROC curve was 0.81, with an optimism estimate of -0.02, indicating absence of optimism. The prediction model of Electronic questionnaire preference in TKR patients had an r^2 of 0.41 with an optimism estimate of -0.24, indicating absence of optimism. The Area under the ROC curve was 0.88, with an optimism estimate of -0.004, indicating absence of optimism.

	Pen and Paper	Electronic	Age and Gender-adjusted Odds Ratio
Mean Follow-up Years (SD)	3.10 (1.09)	3.44 (1.32)	-
Mean Age at Joint Replacement (SD)	70.3 (8.91)	59.9 (9.72)	0.90 (0.86 – 0.93)
% Men:	30,7	45,1	0.61 (0.31 – 1.18)
BMI			
% <25:	17,5	19,6	ref.
% 25-30:	43,3	49	0.86 (0.34 – 2.20)
% 30-35:	23,9	23,5	0.65 (0.23 – 1.83)
% >35:	15,3	7,84	0.41 (0.11 – 1.57)
Indication for Joint Replacement			
% Osteoarthritis:	90,4	81,6	1.29 (0.48 – 3.45)
Patient-reported Charnley Classification			
% A:	14,9	12,2	ref.
% B:	11,2	6,12	0.83 (0.18 – 3.83)
% C:	73,9	81,6	1.35 (0.52 – 3.52)
Highest Completed Level of Schooling			
% University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education:	11,7	37,5	ref.
% Middle Vocational Education and Preparatory Middle Vocational Education:	33,1	50	0.55 (0.24 – 1.26)
% Lower Vocational Education, Elementary Schooling and No Formal Education:	55,2	12,5	0.08 (0.03 – 0.25)
Mean Preoperative SF36 Summary Scores (SD)			
Physical Component Summary	41.1 (8.81)	36.7 (12.9)	1.02 (0.98 – 1.06)
Mental Component Summary	51.2 (10.4)	53.5 (8.87)	1.02 (0.99 – 1.06)

Table 9.4: **Comparison of Patients Who Prefer Pen And Paper Questionnaires With Patients Who Prefer Electronic Questionnaires: Total Knee Replacement.**

Discussion

The vast majority of THR and TKR patients prefer pen-and-paper questionnaires. THR patients who prefer electronic questionnaires are younger, more often male, have completed higher levels of schooling and are more often obese. TKR patients who prefer electronic questionnaires are younger and have completed higher levels of schooling.

A limitation of our study is the mode of questionnaire used to capture the data. In this study, we invited patients to participate by conventional mail. Additionally, all

	Odds Ratio (95%CI)	p-value
Age at Joint Replacement (Years)	0.93 (0.90 – 0.96)	< 0.001
Male vs Female Gender	0.31 (0.17 – 0.56)	< 0.001
BMI: 25-30 vs $<$25	2.06 (1.03 – 4.11)	0.04
BMI: 30-35 vs $<$25	1.17 (0.48 – 2.81)	0.73
BMI: $>$35 vs $<$25	5.49 (1.74 – 17.3)	0.004
Other Indications vs Osteo-Arthritis	0.59 (0.28 – 1.26)	0.17
Charnley Classification: A vs B	0.99 (0.40 – 2.42)	0.98
Charnley Classification: A vs C	0.87 (0.43 – 1.78)	0.70
Schooling: U+HVE+PHVSE vs MVE+PMVE	0.89 (0.45 – 1.77)	0.74
Schooling: U+HVE+PHVSE vs LVE+ES+NFE	0.27 (0.12 – 0.59)	< 0.001
Physical Component Summary	1.00 (0.97 – 1.03)	0.80
Mental Component Summary	1.00 (0.98 – 1.03)	0.87

Table 9.5: **Multivariate Prediction of Electronic Questionnaires Preference: Total Hip Replacement.** Odds ratios > 1 indicate a higher odds of preferring an Electronic Questionnaire, per increasing predictor unit. r^2 = 0.31; AUC = 0.81.
U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education. MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education. LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education.

questionnaires consisted of pen-and-paper questionnaires. Patients willing to participate in this study might be more inclined to prefer pen-and-paper questionnaires than THR and TKR patients in general, thus leading to an overestimation of the proportion of patients preferring pen-and-paper questionnaires. However, we consider it unlikely that the entire preference for pen-and-paper questionnaires is based on such selection bias. Additionally, the identified predictors for electronic questionnaire preference, such as age and completed level of schooling, are plausible, thereby indirectly validating our results.

Strengths of our study include the large sample size, allowing precise estimation and multivariate prediction of patient preference. Although the low r^2 values indicate that not all variance is explained by the predictors, the high AUC values indicate that the prediction models have a high discriminatory ability. The limited extent of optimism in r^2 and AUC estimates indicate that overfitting did not play a role in our study.[231] In other words, it is unlikely that the prediction models in this study have captured the peculiarities in this data set; conversely, it is likely that predictions, based on this data, will be generalisable to other, similar populations.

	Odds Ratio (95%CI)	p-value
Age at Joint Replacement (Years)	0.89 (0.84 – 0.94)	< 0.001
Male Gender	0.53 (0.21 – 1.34)	0.18
BMI: 25-30 vs $<$25	1.05 (0.26 – 4.28)	0.94
BMI: 30-35 vs $<$25	1.27 (0.30 – 5.38)	0.75
BMI: $>$35 vs $<$25	1.59 (0.28 – 8.91)	0.60
Other Indications vs Osteo-Arthritis	2.05 (0.53 – 7.89)	0.30
Charnley Classification: A vs B	1.40 (0.23 – 8.58)	0.72
Charnley Classification: A vs C	2.07 (0.58 – 7.31)	0.26
Schooling: U+HVE+PHVSE vs MVE+PMVE	0.33 (0.13 – 0.85)	0.02
Schooling: U+HVE+PHVSE vs LVE+ES+NFE	0.04 (0.01 – 0.15)	< 0.001
Physical Component Summary	1.01 (0.96 – 1.06)	0.63
Mental Component Summary	0.99 (0.95 – 1.04)	0.75

Table 9.6: **Multivariate Prediction of Electronic Questionnaires Preference: Total Knee Replacement.** Odds ratios > 1 indicate a higher odds of preferring an Electronic Questionnaire, per increasing predictor unit. r^2 = 0.41; AUC = 0.88.
U+HVE+PHVSE: University, Higher Vocational Education and Preparatory Higher Vocational & Scientific Education. MVE+PMVE: Middle Vocational Education and Preparatory Middle Vocational Education. LVE+ES+NFE: Lower Vocational Education, Elementary Schooling and No Formal Education.

Unfortunately, we do not have any information on the availability of Internet access of our patients. Although The Netherlands is rated as one of the most mature internet markets,[232] recent evidence suggests that non-users of the internet are more likely to be elderly,[233] which could explain pen-and-paper questionnaire preference. Practical advantages of electronic questionnaires are stressed in the current orthopaedic literature.[24, 234] Patients are sometimes considered to prefer electronic questionnaires, without any evidence supporting this claim.[234] Although electronic questionnaires certainly appear more efficient, our results reveal limitations in line with the findings of Rolfson et al.[225] Future studies, which only measure PROMs using electronic questionnaires, might suffer from limited generalizability, as elderly and lowly educated patients are less likely to participate. Moreover, selection bias might occur if the association of interest is related to age or social class.

When planning a study in which PROMs will be completed by THR and TKR patients at home, we recommend using pen-and-paper questionnaires, despite their logistic limitations. Such studies should at least provide the option of pen-and-paper questionnaires, in order to prevent selection bias by questionnaire mode.

10

Discussion

Thesis Summary

Implants In chapter 2 (p. 9), we systematically searched and appraised the current literature, regarding the probability of revision surgery at ten years for each individual Total Hip Implant. We compared the study results to the NICE benchmark,[35] and found that 8 out of 34 acetabular cups and 15 out of 32 femoral stems outperform the benchmarks. 16 out of 34 acetabular cups and 6 out of 32 femoral stems performed significantly worse than the NICE benchmark. Most studies were of low methodological quality, the risk of bias is therefore high.

Patient and Surgeon Factors In this thesis, we investigated whether two patient characteristics, namely the patients Socio-Economic Position and the preoperative radiographic severity were predictors of improvement in HRQoL and patient satisfaction after THR/TKR.

In chapter 3 (p. 37), we questioned whether more disadvantaged Socio-Economic Position is associated with an lower improvement in Health-Related Quality of Life (HRQoL) and a lower patient satisfaction after THR/TKR in a multi-center cohort study. We found no differences in HRQoL improvement in THR patients and small, clinically irrelevant differences in HRQoL improvement in some subscales for TKR patients. Additionally, we found no differences in patient satisfaction, both for THR patients and TKR patients.

In conclusion, Socio-Economic Position is no useful patient characteristic to predict HRQoL improvement and patient satisfaction in the Netherlands.

In chapter 4 (p. 55), we assessed whether the pre-operative radiographic OA severity is related to the improvement in HRQoL after THR or TKR, both at the population and individual level. Severe OA patients improved more and had a higher probability of a relevant improvement in physical functioning after both THR and TKR. Patient satisfaction was also higher in severe OA TKR patients. In conclusion, the radiographic OA severity could be a useful patient characteristic to predict HRQoL improvement and patient satisfaction.

Research Methodology

Competing Risks: In chapter 5 (p. 71), we assessed how much bias is introduced in the estimation of the probability of revision surgery, when a crucial assumption of the Kaplan-Meier estimator is violated. Independence of the time to event and the censoring distribution is assumed in the Kaplan-Meier estimator. In the presence of competing events, this assumption does not hold. Using the Kaplan-Meier estimator when competing risks are present, will *always* lead to an overestimation of the cumulative probability in question.

Clinimetrics: In chapter 6 (p. 81), we aimed to summarise minimal clinically important differences (MCIDs) after total hip (THR) or knee replacement (TKR) in health-related quality of life (HRQoL), measured using the Short-Form 36 (SF-36). We also aimed to improve the precision of MCID estimates by means of meta-analysis. Our systematic review of the literature yielded three studies, each describing a distinct study population: primary THR, primary TKR and revision THR. No synthesis of study results can be given. The MCID estimates which we have found were not validated using external criteria and had limited precision. Nonetheless, these are the best known estimates of MCIDs in HRQoL after THR and TKR to date. We therefore advise cautious use of these MCIDs as absolute thresholds.

In chapter 7 (p. 93), we aimed to determine Clinically Important Differences (CIDs) in Health-Related Quality of Life (HRQoL) after Total Hip (THR) or Total Knee Replacement

(TKR), using the Short-Form 36 (SF36). CIDs are more relevant than MCIDs in THR and TKR, as one would expect a substantial improvement in HRQoL after joint replacement, instead of just a minimal improvement. CIDs of Physical Functioning, Role Physical, Bodily Pain and Social Functioning were validated by the validation question "knowing what your hip or knee replacement surgery did for you, would you still have undergone this surgery (yes / no)?". CIDs of all other subscales should be used cautiously, as these were not validated using external criteria.

In chapter 8 (p. 105), we aimed to define Patient Acceptable Symptom State (PASS) thresholds for the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) at mid-term follow-up. Receiver operating characteristic (ROC) curves identified a PASS threshold of 42 points for the OHS after THR and 37 points for the OKS after TKR. THR patients with an OHS $\geq$ 42 points and TKR patients with an OKS $\geq$ 37 points had a higher Numeric Rating Scale of Satisfaction and a larger odds of being willing to undergo surgery again. However, PASS thresholds differed considerably between relevant subgroups. PASS thresholds appear larger at mid-term follow-up than at 6 months after surgery. Without external validation, we would advise against using these PASS thresholds as absolute thresholds in defining whether or not a patients has attained an acceptable symptom state after THR/TKR.

Questionnaire Mode Preference: In chapter 9 (p. 119), we assessed patient preference for the questionnaire mode in a multi-center cohort study. The majority of THR and TKR patients prefer pen-and-paper questionnaires. Patients preferring electronic questionnaires differed from patients who preferred pen-and-paper questionnaires. Restricting the mode of PROMs to electronic questionnaires might introduce selection bias.

Recommendations for Future Research

Implants The ideal THI has a low probability of revision surgery and has little systemic adverse effects. In chapter 2 (p. 9), we systematically searched and appraised the literature, regarding the probability of revision surgery of THI at ten years follow-up. A

number of THI outperform current benchmarks. However, the majority of found survival estimates were based on a single study, performed in a single center. The results of this study should not be viewed as conclusive evidence, but as the best available evidence at *this* point in time.

Post-market surveillance studies, which are summarised in our systematic review (chapter 2 (p. 9)), remain of vital importance to detect implants with a high probability of revision surgery. A major drawback of such studies is that, if published at all, their results are available many years after an implant has been introduced in clinical practice. National joint registries, given an annual update, solve the first problem, but not the second. Solely relying on post-market surveillance and national joint registry studies to detect poor implants will therefore expose many patients to unproven designs and facilitates large-scale implant recalls, as we have seen in the ASR case.[111]

Imaging techniques, such as 3D Röntgen Stereophotogrammetry Analysis (RSA), could play a crucial role in preventing future implant disasters. The probability of revision surgery can be predicted using RSA.[42, 114, 235] A recent study shows that implants, which have published RSA studies at two years follow-up, have 22–35% less revisions up to 5 years after surgery.[119] Phased introduction of new implants using RSA could therefore lead to better patient care and could substantially reduce health-care costs associated with revision surgery.

Patient and Surgeon Factors We have assessed the role of the patients Socio-Economic Position and the severity of preoperative radiographic OA in predicting the Patient-Reported Outcome after THR and TKR. The number of potential predictors of the improvement in clinical outcome is endless.

One potential predictor, which is often stressed in the literature, is the preoperative patient expectation of the outcome after joint replacement.[166] These expectations can be modified by preoperative educational classes, which opens up possibilities for preoperative optimalisation.[236] In future cohort studies, the role of pre-operative expectations on the probability of a relevant improvement in HRQoL should be studied, in order to investigate whether or not expectation management can lower the relatively

high rate of dissatisfaction after joint replacement.

Research Methodology

Competing Risks: In chapter 5 (p. 71), we have shown that the Kaplan-Meier (KM) estimator introduces bias in the presence of competing events. When estimating the cumulative probability of revision surgery, competing events are likely to occur in the case of THR or TKR. A recently developed guideline for the statistical analysis of arthroplasty data acknowledges that the KM estimator yields biased results.[237] Unfortunately, the authors of this guideline miss the point in interpreting the consequences of this bias. Two poor arguments in favour of the KM estimator are proposed. In the first place, "Is the difference (i.e. the amount of bias)... clinically important?" We have shown that the amount of bias depends on the number of competing events (i.e. the number of patients who have died), compared to the number of events of interest (i.e. the number of patients who have undergone revision surgery). However, why would one be willing to accept *any* form of bias, especially when it is possible to eliminate such bias using freely available tools? In the second place, the authors state that "The KM estimates of implant failure are more clinically meaningful and straightforward to interpret for clinicians and patients". This argumentation is flawed, since both the KM estimator and the cause-specific cumulative incidence estimator estimate the cumulative probability of a certain event as a function of time. Thus, the clinical meaning of the KM estimator and the cumulative incidence estimator is identical. The only difference is that the cumulative probability of being event-free is presented by the KM estimator, while the cumulative incidence estimator presents the cumulative probability of having the event of interest. Surely, getting accustomed to this slightly different way of presenting the probability of revision surgery is worthwhile, as it permits unbiased estimation of the outcome of interest.

Clinimetrics: In this thesis we have summarised the literature regarding MCIDs in HRQoL after THR and TKR, we have estimated CIDs in HRQoL after THR and TKR using an innovative approach and we have estimated PASS in Joint-specific Patient Reported Outcome Measures.

A number of issues remain to be addressed. To date, no CIDs have been established for neither OHS nor OKS, two often used joint specific Patient Reported Outcome Measures. In esstimating these CIDs, future studies could compare our innovative approach to the approach of Chesworth et al.[203]

Questionnaire Mode Preference: In chapter 9 (p. 119), we found that the vast majority of THR and TKR patients prefer pen-and-paper questionnaires, when participating in a cohort study on the improvement in Health-Related Quality of Life after THR or TKR.

In the past few years, tablet computers have gained in popularity. Recent evidence suggests that the acceptance and satisfaction rates of tablet computers are high amongst senior users.[238] It would be interesting to see whether or not the rising popularity of tablet computers will affect the patient preference for pen-and-paper questionnaires in the near future.

Bibliography

1 Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am*. 2004;86-A(5):963–74.

2 Havelin LI, Fenstad AM, Salomonsson R, Mehnert F, Furnes O, Overgaard Sr, Pedersen AB, Herberts P, Kärrholm J, Garellick G. The Nordic Arthroplasty Register Association: a unique collaboration between 3 national hip arthroplasty registries with 280,201 THRs. *Acta Orthop*. 2009;80(4):393–401.

3 Robertsson O, Bizjajeva S, Fenstad AM, Furnes O, Lidgren L, Mehnert F, Odgaard A, Pedersen AB, Havelin LI. Knee arthroplasty in Denmark, Norway and Sweden. *Acta Orthop*. 2010;81(1):82–9.

4 Chang RW, Pellisier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *JAMA*. 1996;275(11):858–65.

5 O'Shea K, Bale E, Murray P. Cost analysis of primary total hip replacement. *Ir Med J*. 2002;95(6):177–180.

6 Räsänen P, Paavolainen P, Sintonen H, Koivisto AM, Blom M, Ryynänen OP, Roine RP. Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthop*. 2007;78(1):108–115.

7 Hamilton DF, Clement ND, Burnett R, Patton JT, Moran M, Howie CR, Simpson AHRW, Gaston P. Do modern total knee replacements offer better value for money? A health economic analysis. *Int Orthop*. 2013;37(11):2147–52.

8 LROI. Annual reports. Insight into quality of orthopaedic care in the Netherlands. URL `http://www.lroi.nl/en/annual-reports`

9 Otten R, van Roermund PM, Picavet HSJ. [Trends in the number of knee and hip arthroplasties: considerably more knee and hip prostheses due to osteoarthritis in 2030]. *Ned Tijdschr Geneeskd*. 2010;154:A1534.

10 Advice by the Board of the NOV on the use of primary hip prostheses. URL `http://www.orthopeden.org/`

11 Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ open*. 2012;2(1):e000435.

12 Gandhi R, Davey JR, Mahomed NN. Predicting patient dissatisfaction following joint replacement surgery. *J Rheumatol*. 2008;35(12):2415–8.

13 Noble PC, Conditt MA, Cook KF, Mathis KB. The John Insall Award: Patient expectations affect satisfaction with total knee arthroplasty. *Clin Orthop*. 2006;452:35–43.

14 Brokelman RBG, van Loon CJM, Rijnberg WJ. Patient versus surgeon satisfaction after total hip arthroplasty. *J Bone Joint Surg Br*. 2003;85-B(4):495–498.

15 Nilsdotter AK, Toksvig-Larsen S, Roos EM. Knee arthroplasty: are patients' expectations fulfilled? A prospective study of pain and function in 102 patients with 5-year follow-up. *Acta Orthop*. 2009;80(1):55–61.

16 Lingard EA, Sledge CB, Learmonth ID. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. *J Bone Joint Surg Am*. 2006;88(6):1201–1207.

17 Garellick G, Malchau H, Herberts P. Survival of hip replacements. A comparison of a randomized trial and a registry. *Clin Orthop Relat Res*. 2000;(375):157–167.

18 Baker PN, van der Meulen JH, Lewsey J, Gregg PJ. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. *J Bone Joint Surg Br*. 2007;89(7):893–900.

19 Mancuso CA, Salvati EA, Johanson NA, Peterson MG, Charlson ME. Patients' expectations and satisfaction with total hip arthroplasty. *J Arthroplasty*. 1997;12(4):387–96.

20 Vissers MM, de Groot IB, Reijman M, Bussmann JB, Stam HJ, Verhaar JA. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. *BMC Musculoskelet Disord*. 2010;11:121.

21 Keurentjes JC, Blane D, Bartley M, Keurentjes JJB, Fiocco M, Nelissen RGR. Socio-Economic Position has no effect on Improvement in Health-Related Quality of Life and Patient Satisfaction in Total Hip and Knee Replacement: A Cohort Study. *PLoS ONE*. 2013;8(e56785):e56785.

22 Kaplan E, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958;53(282):457–481.

23 Van Gelder M, Pijpe A. E - epidemiology: a comprehensive update. *OA Epidemiology*. 2013;1(2):1.

24 Saleh KJ, Radosevich DM, Kassim RA, Moussa M, Dykes D, Bottolfson H, Gioe TJ, Robinson H. Comparison of commonly used orthopaedic outcome measures using palm-top computers and paper surveys. *J Orthop Res*. 2002;20:1146–1151.

25 So-Osman C, Nelissen R, Te Slaa R, Coene L, Brand R, Brand A. A randomized comparison of transfusion triggers in elective orthopaedic surgery using leucocyte-depleted red blood cells. *Vox Sang*. 2010;98(1):56–64.

26 So-Osman C, Nelissen R, Brand R, Faber F, Slaa RT, Stiggelbout A, Brand A. The impact of a restrictive transfusion trigger on post-operative complication rate and well-being following elective orthopaedic surgery: a post-hoc analysis of a randomised study. *Blood Transfus*. 2013;11(2):289–95.

27 Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AWMM, Pöll RG, Kroon HM, Vliet Vlieland TPM, Nelissen RG. Patients with Severe Radiographic Osteoarthritis Have a Better Prognosis in Physical Functioning after Hip and Knee Replacement: A Cohort-Study. *PLoS ONE*. 2013;8(4):e59500.

28 Keurentjes JC, Fiocco M, Nelissen RG. Willingness to undergo surgery again validated clinically important differences in health-related quality of life after total hip replacement or total knee replacement surgery. *J Clin Epidemiol*. 2014;67(1):114–20.

29 Murray DW, Carr AJ, Bulstrode CJ. Which primary total hip replacement? *J Bone Joint Surg Br*. 1996;78(3):510.

30 Bozic KJ, Durbhakula S, Berry DJ, Naessens JM, Rappaport K, Cisternas M, Saleh KJ, Rubash HE. Differences in patient and procedure characteristics and hospital resource use in primary and revision total joint arthroplasty: a multicenter study. *J Arthroplasty*. 2005;20(7 Suppl 3):17–25.

31 Bozic KJ, Katz P, Cisternas M, Ono L, Ries MD, Showstack J. Hospital resource utilization for primary and revision total hip arthroplasty. *J Bone Joint Surg Am*. 2005;87(3):570–576.

32 Ong KL, Mowat FS, Chan N, Lau E, Halpern MT, Kurtz SM. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. *Clin Orthop*. 2006;446(446):22–28.

33 Vanhegan IS, Malik AK, Jayakumar P, Ul Islam S, Haddad FS. A financial analysis of revision hip arthroplasty: the economic burden in relation to the national tariff. *J Bone Joint Surg Br*. 2012;94(5):619–23.

34 Schreurs BW, Keurentjes JC, Gardeniers JWM, Verdonschot N, Slooff TJJH, Veth RPH. Acetabular revision with impacted morsellised cancellous bone grafting and a cemented acetabular component: a 20- to 25-year follow-up. *J Bone Joint Surg Br*. 2009;91(9):1148–53.

35 National Institute for Health and Care Excellence: Review of TA2 hip disease replacement prostheses and TA44 hip disease metal on metal resurfacing.
URL http://publications.nice.org.uk/

36 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535–b2535.

37 Pijls BG, Dekkers OM, Middeldorp S, Valstar ER, van der Heide HJL, Van der Linden-Van der Zwaag HMJ, Nelissen RGHH. AQUILA: assessment of quality in lower limb arthroplasty. An expert Delphi consensus for total knee and total hip arthroplasty. *BMC Musculoskelet Disord*. 2011;12:173.

38 Thiessen Philbrook H, Barrowman N, Garg AX. Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *J Clin Epidemiol*. 2007;60(3):228–40.

39 Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol*. 2006;59(1):7–10.

40 Fiocco M, Putter H, van Houwelingen JC. Meta-analysis of pairs of survival curves under heterogeneity: a Poisson correlated gamma-frailty approach. *Stat Med*. 2009; 28(30):3782–97.

41 Putter H, Fiocco M, Stijnen T. Meta-analysis of diagnostic test accuracy studies with multiple thresholds using survival methods. *Biom J*. 2010;52(1):95–110.

42 Pijls BG, Valstar ER, Nouta KA, Plevier JW, Fiocco M, Middeldorp S, Nelissen RG. Early migration of tibial components is associated with late revision: a systematic review and meta-analysis of 21,000 knee arthroplasties. *Acta Orthop*. 2012;83(6):614–24.

43 *R: A Language and Environment for Statistical Computing*.
URL http://www.r-project.org/

44 Svenska Höftprotesregistret.
URL http://www.shpr.se/en/Publications/DocumentsReports.aspx

45 Eskelinen A, Remes V, Helenius I, Pulkkinen P, Nevalainen J, Paavolainen P. Uncemented total hip arthroplasty for primary osteoarthritis in young patients: a mid-to long-term follow-up study from the Finnish Arthroplasty Register. *Acta Orthop*. 2006;77(1):57–70.

46 Rogers A, Kulkarni R, Downes EM. The ABG hydroxyapatite-coated hip prosthesis: one hundred consecutive operations with average 6-year follow-up. *J Arthroplasty*. 2003; 18(5):619–25.

47 Hamilton WG, Calendine CL, Beykirch SE, Hopper RH, Engh CA. Acetabular fixation options: first-generation modular cup curtain calls and caveats. *J Arthroplasty*. 2007; 22(4 Suppl 1):75–81.

48 Philippe M, Ameil M. Survival analysis at 10 years of a cohort of 297 atlas total hip prostheses. *Eur J Orthop Surg Traumatol*. 2007;17(6):573–578.

49 Ritter MA. The cemented acetabular component of a total hip replacement. All polyethylene versus metal backing. *Clin Orthop*. 1995;(311):69–75.

50 Amstutz HC, Le Duff MJ, Campbell PA, Gruen TA, Wisk LE. Clinical and radiographic results of metal-on-metal hip resurfacing with a minimum ten-year follow-up. *J Bone Joint Surg Am*. 2010;92(16):2663–71.

51 Veitch SW, Whitehouse SL, Howell JR, Hubble MJW, Gie GA, Timperley AJ. The concentric all-polyethylene Exeter acetabular component in primary total hip replacement. *J Bone Joint Surg Br*. 2010;92(10):1351–5.

52 Vassan UT, Sharma S, Chowdary KP, Bhamra MS. Uncemented metal-on-metal acetabular component: follow-up of 112 hips for a minimum of 5 years. *Acta Orthop*. 2007;78(4):470–8.

53 Keisu K, Lindgren JU. Outcome of total hip arthroplasties with cement. Results after implantation of the Harris Design 2 prosthesis, 1983-1985. *J Arthroplasty*. 1996; 11(7):789–96.

54 Petersen MB, Poulsen IH, Thomsen J, Solgaard S. The hemispherical Harris-Galante acetabular cup, inserted without cement. The results of an eight to eleven-year follow-up of one hundred and sixty-eight hips. *J Bone Joint Surg Am*. 1999;81(2):219–24.

55 Metzner G, Strassl R, Neumann DRP, Hitzl W, Ramsauer T, Dorn U. [Long-term results of the cementless Hofer-Imhof threaded titanium cup]. *Z Orthop Unfall*;147(2):166–74.

56 Shetty AA, Slack R, Tindall A, James KD, Rand C. Results of a hydroxyapatite-coated (Furlong) total hip replacement: a 13- to 15-year follow-up. *J Bone Joint Surg Br*. 2005; 87(8):1050–4.

57 Partio E, von Bonsdorff H, Wirta J, Avikainen V. Survival of the Lubinus hip prosthesis. An eight- to 12-year follow-up evaluation of 444 cases. *Clin Orthop*. 1994;(303):140–6.

58 Bourne RB, Rorabeck CH, Patterson JJ, Guerin J. Tapered titanium cementless total hip replacements: a 10- to 13-year followup study. *Clin Orthop*. 2001;(393):112–20.

59 Berli BJ, Schäfer D, Morscher EW. Ten-year survival of the MS-30 matt-surfaced cemented stem. *J Bone Joint Surg Br*. 2005;87(7):928–33.

60 Gwynne-Jones DP, Garneti N, Wainwright C, Matheson JA, King R. The Morscher Press Fit acetabular component: a nine- to 13-year review. *J Bone Joint Surg Br*. 2009; 91(7):859–64.

61 Farizon F, de Lavison R, Azoulai JJ, Bousquet G. Results with a cementless alumina-coated cup with dual mobility. A twelve-year follow-up study. *Int Orthop*. 1998;22(4):219–24.

62 Emms NW, Stockley I, Hamer AJ, Wilkinson JM. Long-term outcome of a cementless, hemispherical, press-fit acetabular component: survivorship analysis and dose-response relationship to linear polyethylene wear. *J Bone Joint Surg Br*. 2010;92(6):856–61.

63 Diks M, van den Broek C, Anderson P, van Limbeek J, Spruit M. The uncemented, titanium-coated rm cup: Survival and analyses of failures. *Hip Int*;(Vol. 15 Issue 2 (April - June 2005)):71–77.

64 Levy BA, Berry DJ, Pagnano MW. Long-term survivorship of cemented all-polyethylene acetabular components in patients > 75 years of age. *J Arthroplasty*. 2000;15(4):461–7.

65 Keisu KS, Orozco F, Sharkey PF, Hozack WJ, Rothman RH, McGuigan FX. Primary cementless total hip arthroplasty in octogenarians. Two to eleven-year follow-up. *J Bone Joint Surg Am*. 2001;83-A(3):359–63.

66 Suckel A, Geiger F, Kinzl L, Wulker N, Garbrecht M. Long-term results for the uncemented Zweymuller/Alloclassic hip endoprosthesis. A 15-year minimum follow-up of 320 hip operations. *J Arthroplasty*. 2009;24(6):846–53.

67 Dohle J, Becker W, Braun M. [Radiological analysis of osseointegration after implantation of the Zweymüller-Alloclassic total hip system]. *Z Orthop Unfall*;139(6):517–24.

68 Ochs U, Eingartner C, Volkmann R, Ochs BG, Huber C, Weller S, Weise K. Prospective long-term follow-up of the cementless bicontact hip stem with plasmapore coating. *Z Orthop Unfall*;145 Suppl:S3–8.

69 Bodén H, Salemyr M, Sköldenberg O, Ahl T, Adolphson P. Total hip arthroplasty with an uncemented hydroxyapatite-coated tapered titanium stem: results at a minimum of 10 years' follow-up in 104 hips. *J Orthop Sci*. 2006;11(2):175–9.

70 Meding JB, Keating EM, Ritter MA, Faris PM, Berend ME. Minimum ten-year follow-up of a straight-stemmed, plasma-sprayed, titanium-alloy, uncemented femoral component in primary total hip arthroplasty. *J Bone Joint Surg Am*. 2004;86-A(1):92–7.

71 Rasquinha VJ, Ranawat CS. Durability of the cemented femoral stem in patients 60 to 80 years old. *Clin Orthop Relat Res*. 2004;419:115–123.

72 Chatelet JC, Setiey L. [Long term bone behavior in total primary hip arthroplasty with a fully hydroxyapatite-coated femoral stem: a continuous series of 120 cases with twelve years follow-up]. *Orthop Traumatol Surg Res*. 2004;90(7):628–35.

73 Young L, Duckett S, Dunn A. The use of the cemented Exeter Universal femoral stem in a District General Hospital: a minimum ten-year follow-up. *J Bone Joint Surg Br*. 2009; 91(2):170–5.

74 Mannan K, Freeman MAR, Scott G. The Freeman femoral component with hydroxyapatite coating and retention of the neck: an update with a minimum follow-up of 17 years. *J Bone Joint Surg Br*. 2010;92(4):480–5.

75 Datir SP, Kurta IC, Wynn-Jones CH. Ten-year survivorship of rough-surfaced femoral stem with geometry similar to Charnley femoral stem. *J Arthroplasty*. 2006;21(3):392–7.

76 Jewett BA, Collis DK. Radiographic failure patterns of polished cemented stems. *Clin Orthop*. 2006;453:132–6.

77 Grant P, Nordsletten L. Total hip arthroplasty with the Lord prosthesis. A long-term follow-up study. *J Bone Joint Surg Am*. 2004;86-A(12):2636–41.

78 Savilahti S, Myllyneva I, Pajamäki KJ, Lindholm TS. Survival of Lubinus straight (IP) and curved (SP) total hip prostheses in 543 patients after 4-13 years. *Arch Orthop Trauma Surg*. 1997;116(1-2):10–3.

79 Emerson RH, Head WC, Emerson CB, Rosenfeldt W, Higgins LL. A comparison of cemented and cementless titanium femoral components used for primary total hip arthroplasty: a radiographic and survivorship study. *J Arthroplasty*. 2002;17(5):584–91.

80 Lombardi AV, Berend KR, Mallory TH, Skeels MD, Adams JB. Survivorship of 2000 tapered titanium porous plasma-sprayed femoral components. *Clin Orthop*. 2009; 467(1):146–54.

81 Capello WN, D'Antonio JA, Jaffe WL, Geesink RG, Manley MT, Feinberg JR. Hydroxyapatite-coated femoral components: 15-year minimum followup. *Clin Orthop*. 2006;453:75–80.

82 Mohajer MA, Hofmann F, Graf R. Retrospective and consecutive analysis of the long-term outcomes of the SBG stem: a 15-year follow-up study. *Arch Orthop Trauma Surg*. 2010; 130(2):185–90.

83 Santori FS, Santori N. Mid-term results of a custom-made short proximal loading femoral component. *J Bone Joint Surg Br*. 2010;92(9):1231–7.

84 Parvizi J, Keisu KS, Hozack WJ, Sharkey PF, Rothman RH. Primary total hip arthroplasty with an uncemented femoral component: a long-term study of the Taperloc stem. *J Arthroplasty*. 2004;19(2):151–6.

85 Johnsson R, Thorngren KG, Persson BM. Revision of total hip replacement for primary osteoarthritis. *J Bone Joint Surg Br*. 1988;70(1):56–62.

86 Espehaug B, Furnes O, Havelin LI, Engesaeter LB, Vollset SE. The type of cement and failure of total hip replacements. *J Bone Joint Surg Br*. 2002;84(6):832–8.

87 Espehaug B, Furnes O, Engesaeter LB, Havelin LI. 18 years of results with cemented primary hip prostheses in the Norwegian Arthroplasty Register: concerns about some newer implants. *Acta Orthop*. 2009;80(4):402–412.

88 Hop JD, Callaghan JJ, Olejniczak JP, Pedersen DR, Brown TD, Johnston RC. The Frank Stinchfield Award. Contribution of cable debris generation to accelerated polyethylene wear. *Clin Orthop*. 1997;(344):20–32.

89 Skutek M, Bourne RB, Rorabeck CH, Burns A, Kearns S, Krishna G. The twenty to twenty-five-year outcomes of the Harris design-2 matte-finished cemented total hip replacement. A concise follow-up of a previous report. *J Bone Joint Surg Am*. 2007; 89(4):814–8.

90 Neumann DRP, Thaler C, Hitzl W, Huber M, Hofstädter T, Dorn U. Long-term results of a contemporary metal-on-metal total hip arthroplasty: a 10-year follow-up study. *J Arthroplasty*. 2010;25(5):700–8.

91 Macheras G, Kateros K, Kostakos A, Koutsostathis S, Danomaras D, Papagelopoulos PJ. Eight- to ten-year clinical and radiographic outcome of a porous tantalum monoblock acetabular component. *J Arthroplasty*. 2009;24(5):705–9.

92 de Jong PT, van der Vis HM, de Man FHR, Marti RK. Weber rotation total hip replacement: a prospective 5- to 20-year followup study. *Clin Orthop*. 2004;(419):107–14.

93 Zwartele R, Peters A, Brouwers J, Olsthoorn P, Brand R, Doets C. Long-term results of cementless primary total hip arthroplasty with a threaded cup and a tapered, rectangular titanium stem in rheumatoid arthritis and osteoarthritis. *Int Orthop*. 2008;32(5):581–7.

94 Delaunay C, Kapandji AI. Survival analysis of cementless grit-blasted titanium total hip arthroplasties. *J Bone Joint Surg Br*. 2001;83(3):408–13.

95 Kobayashi S, Saito N, Horiuchi H, Iorio R, Takaoka K. Poor bone quality or hip structure as risk factors affecting survival of total-hip arthroplasty. *Lancet*. 2000;355(9214):1499–504.

96 von Schewelov T, Sanzén L, Besjakov J, Carlsson A. The Elite-Plus stem migrates more than the flanged Charnley stem. *Acta Orthop*. 2010;81(3):280–5.

97 Mäkelä K, Eskelinen A, Pulkkinen P, Paavolainen P, Remes V. Cemented total hip replacement for primary osteoarthritis in patients aged 55 years or older: results of the 12 most common cemented implants followed for 25 years in the Finnish Arthroplasty Register. *J Bone Joint Surg Br*. 2008;90(12):1562–9.

98 Hook S, Moulder E, Yates PJ, Burston BJ, Whitley E, Bannister GC. The Exeter Universal stem: a minimum ten-year review from an independent centre. *J Bone Joint Surg Br*. 2006;88(12):1584–90.

99 Skinner JA, Todo S, Taylor M, Wang JS, Pinskerova V, Scott G. Should the cement mantle around the femoral component be thick or thin? *J Bone Joint Surg Br*. 2003; 85(1):45–51.

100 Journeaux SF, Morgan DA, Donnelly WJ. The medium-term results of a cemented Freeman femoral neck-retaining prosthesis. *J Bone Joint Surg Br*. 2000;82(2):188–91.

101 Skinner JA, Kroon PO, Todo S, Scott G. A femoral component with proximal HA coating. An analysis of survival and fixation at up to ten years. *J Bone Joint Surg Br*. 2003; 85(3):366–70.

102 Räber DA, Czaja S, Morscher EW. Fifteen-year results of the Müller CoCrNiMo straight stem. *Arch Orthop Trauma Surg*. 2001;121(1-2):38–42.

103 Riede U, Lüem M, Ilchmann T, Eucker M, Ochsner PE. The M.E Müller straight stem prosthesis: 15 year follow-up. Survivorship and clinical results. *Arch Orthop Trauma Surg*. 2007;127(7):587–92.

104 Baumann B, Hendrich C, Barthel T, Bockholt M, Walther M, Eulert J, Rader CP. 9- to 11-year results of cemented titanium mueller straight stem in total hip arthroplasty. *Orthopedics*. 2007;30(7):551–7.

105 Kovac S, Trebse R, Milosev I, Pavlovcic V, Pisot V. Long-term survival of a cemented titanium-aluminium-vanadium alloy straight-stem femoral component. *J Bone Joint Surg Br*. 2006;88(12):1567–73.

106 Jaffe WL, Hawkins CA. Normalized and proportionalized cemented femoral stem survivorship at 15 years. *J Arthroplasty*. 1999;14(6):708–13.

107 Engesaeter LB, Lie SA, Espehaug B, Furnes O, Vollset SE, Havelin LI. Antibiotic prophylaxis in total hip arthroplasty: effects of antibiotic prophylaxis systemically and in bone cement on the revision rate of 22,170 primary hip replacements followed 0-14 years in the Norwegian Arthroplasty Register. *Acta Orthop Scand*. 2003;74(6):644–51.

108 Thanner J, Freij-Larsson C, Kärrholm J, Malchau H, Wesslén B. Evaluation of Boneloc. Chemical and mechanical properties, and a randomized clinical study of 30 total hip arthroplasties. *Acta Orthop Scand*. 1995;66(3):207–14.

109 Muirhead-Allwood SK. Lessons of a hip failure. *BMJ*. 1998;316(7132):644.

110 Faro LM, Huiskes R. Quality assurance of joint replacement. Legal regulation and medical judgement. *Acta Orthop Scand Suppl*. 1992;250:1–33.

111 Cohen D. Out of joint: the story of the ASR. *BMJ*. 2011;342:d2905.

112 Franklin J, Robertsson O, Gestsson J, Lohmander LS, Ingvarsson T. Revision and complication rates in 654 Exeter total hip replacements, with a maximum follow-up of 20 years. *BMC Musculoskelet Disord*. 2003;4:6.

113 Sundfeldt M, Carlsson LV, Johansson CB, Thomsen P, Gretzer C. Aseptic loosening, not only a question of wear: a review of different theories. *Acta Orthop*. 2006;77(2):177–197.

114 Valstar ER, Nelissen RG, Reiber JH, Rozing PM. The use of Roentgen stereophotogrammetry to study micromotion of orthopaedic implants. *ISPRS J Photogramm Remote Sens*. 2002;56(5-6):376–389.

115 Nelissen RG, Brand R, Rozing PM. Survivorship analysis in total condylar knee arthroplasty. A statistical review. *J Bone Joint Surg Am*. 1992;74(3):383–389.

116 Keurentjes JC, Fiocco M, Schreurs BW, Pijls BG, Nouta KA, Nelissen RGHH. Revision surgery is overestimated in hip replacement. *Bone Joint Res*. 2012;1(10):258–262.

117 Labek G, Sekyra K, Pawelka W, Janda W, Stöckl B. Outcome and reproducibility of data concerning the Oxford unicompartmental knee arthroplasty: a structured literature review including arthroplasty registry data. *Acta Orthop*. 2011;82(2):131–5.

118 Wilmshurst P. The regulation of medical devices. *BMJ*. 2011;342:d2822.

119 Nelissen RGHH, Pijls BG, Kärrholm J, Malchau H, Nieuwenhuijse MJ, Valstar ER. RSA and registries: the quest for phased introduction of new implants. *J Bone Joint Surg Am*. 2011;93 Suppl 3:62–5.

120 Sedrakyan A. Hip resurfacing: a complex challenge for device regulation. *Lancet*. 2012; 380(9855):1720–2.

121 McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J, Aronson JK, Barkun JS, Blazeby JM, Boutron IC, Clavien PA, Cook JA, Ergina PL, Feldman LS, Maddern GJ, Reeves BC, Seiler CM, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Meakins J, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374(9695):1105–12.

122 McCulloch P, Barkun J, Sedrakyan A. Implantable device regulation in Europe. *Lancet*. 2012;380(9843):729.

123 Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am*. 2007; 89(4):780–785.

124 Krieger N. A glossary for social epidemiology. *J Epidemiol Community Health*. 2001; 55(10):693–700.

125 Lostao L, Regidor E, Gimeno D, Netuveli G, Blane D. Socioeconomic patterns in health services use in Great Britain and Spain before and after the health system reforms of the 1990s. *Health Place*. 2011;17(3):830–5.

126 Quintana JM, Escobar A, Aguirre U, Lafuente I, Arenaza JC. Predictors of health-related quality-of-life change after total hip arthroplasty. *Clin Orthop*. 2009;467(11):2886–94.

127 Escobar A, Quintana JM, Bilbao A, Azkárate J, Güenaga JI, Arenaza JC, Gutierrez LF. Effect of patient characteristics on reported outcomes after total knee replacement. *Rheumatology*. 2007;46(1):112–9.

128 Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA*;272(8):619–26.

129 Muldoon MF, Barger SD, Flory JD, Manuck SB. What are quality of life measurements measuring? *BMJ*. 1998;316(7130):542–5.

130 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–83.

131 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care*. 1988;26(7):724–35.

132 Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ*. 1993;306(6890):1437–40.

133 Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrips E. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol*. 1998;51(11):1055–68.

134 Ostendorf M, van Stel HF, Buskens E, Schrijvers AJP, Marting LN, Verbout AJ, Dhert WJA. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br*. 2004;86(6):801–808.

135 Grosse Frie K, van der Meulen J, Black N. Single item on patients' satisfaction with condition provided additional insight into impact of surgery. *J Clin Epidemiol*. 2012; 65(6):619–26.

136 Charnley J. The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. *J Bone Joint Surg Br*. 1972;54(1):61–76.

137 Dunbar MJ, Robertsson O, Ryd L. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. *Acta Orthop Scand*. 2004; 75(2):119–26.

138 Dempster AP, Laird NM, Rubin DB. Maximum Likelihood Estimation from Incomplete Data Via the EM Algorithm. *J R Stat Soc Series B Stat Methodol*. 1977;39(1):1–38.

139 Honaker J, King G, Blackwell M. AMELIA II : A Program for Missing Data. *J Stat Softw*. 2010;45(7):1–53.

140 King G, Honaker J, Joseph A, Scheve K. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *Am Polit Sci Rev*. 2001;95(1):49–69.

141 Keurentjes JC, Van Tol FR, Fiocco M, Schoones JW, Nelissen RG. Minimal clinically important differences in health-related quality of life after total hip or knee replacement. *Bone Joint Res*. 2012;1(5):71–77.

142 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, vol. 519. Wiley. 1987.

143 Malchau H, Garellick G, Eisler T, Kärrholm J, Herberts P. Presidential guest address: the Swedish Hip Registry: increasing the sensitivity by patient outcome data. *Clin Orthop*. 2005;441:19–29.

144 Hartge P. Participation in population studies. *Epidemiology*. 2006;17(3):252–4.

145 Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol*. 2007; 17(9):643–53.

146 Jones J. The effects of non-response on statistical inference. *J Health Soc Policy*. 1996; 8(1):49–62.

147 Bruyère O, Ethgen O, Neuprez A, Zégels B, Gillet P, Huskin JP, Reginster JY. Health-related quality of life after total knee or hip replacement for osteoarthritis: a 7-year prospective study. *Arch Orthop Trauma Surg*. 2012;132(11):1583–7.

148 Ng CY, Ballantyne JA, Brenkel IJ. Quality of life and functional outcome after primary total hip replacement. A five-year follow-up. *J Bone Joint Surg Br*. 2007;89(7):868–73.

149 Vissers MM, Bussmann JB, Verhaar JAN, Arends LR, Furlan AD, Reijman M. Recovery of physical functioning after total hip arthroplasty: systematic review and meta-analysis of the literature. *Phys Ther*. 2011;91(5):615–29.

150 Allen Butler R, Rosenzweig S, Myers L, Barrack RL. The Frank Stinchfield Award: the impact of socioeconomic factors on outcome after THA: a prospective, randomized study. *Clin Orthop*. 2011;469(2):339–47.

151 Davis ET, Lingard EA, Schemitsch EH, Waddell JP. Effects of socioeconomic status on patients' outcome after total knee arthroplasty. *Int J Qual Health Care*. 2008;20(1):40–6.

152 Mujica-Mota RE, Tarricone R, Ciani O, Bridges JF, Drummond M. Determinants of demand for total hip and knee arthroplasty: a systematic literature review. *BMC Health Serv Res*. 2012;12(1):225.

153 Agabiti N, Picciotto S, Cesaroni G, Bisanti L, Forastiere F, Onorati R, Pacelli B, Pandolfi P, Russo A, Spadea T, Perucci Ca. The influence of socioeconomic status on utilization and outcomes of elective total hip replacement: a multicity population-based longitudinal study. *Int J Qual Health Care*. 2007;19(1):37–44.

154 Brennan SL, Stanford T, Wluka AE, Henry MJ, Page RS, Graves SE, Kotowicz Ma, Nicholson GC, Pasco Ja. Cross-sectional analysis of association between socioeconomic status and utilization of primary total hip joint replacements 2006-7: Australian Orthopaedic Association National Joint Replacement Registry. *BMC Musculoskelet Disord*. 2012;13(1):63.

155 Dixon T, Shaw M, Ebrahim S, Dieppe P. Trends in hip and knee joint replacement: socioeconomic inequalities and projections of need. *Ann Rheum Dis*. 2004;63(7):825–30.

156 Hawkins K, Escoto KH, Ozminkowski RJ, Bhattarai GR, Migliori RJ, Yeh CS. Disparities in major joint replacement surgery among adults with Medicare supplement insurance. *Popul Health Manag*. 2011;14(5):231–8.

157 Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Geographical variation in the provision of elective primary hip and knee replacement: the role of socio-demographic, hospital and distance variables. *J Public Health*. 2009;31(3):413–22.

158 Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Equity in access to total joint replacement of the hip and knee in England: cross sectional study. *BMJ*. 2010;341(aug11 1):c4092–c4092.

159 Majeed A, Eliahoo J, Bardsley M, Morgan D, Bindman AB. Variation in coronary artery bypass grafting, angioplasty, cataract surgery, and hip replacement rates among primary care groups in London: association with population and practice characteristics. *J Public Health Med*. 2002;24(1):21–6.

160 Peterson MG, Hollenberg JP, Szatrowski TP, Johanson NA, Mancuso CA, Charlson ME. Geographic variations in the rates of elective total hip and knee arthroplasties among Medicare beneficiaries in the United States. *J Bone Joint Surg Am*. 1992;74(10):1530–9.

161 Steel N, Melzer D, Gardener E, McWilliams B. Need for and receipt of hip and knee replacement–a national population survey. *Rheumatology*. 2006;45(11):1437–41.

162 Yong PFK, Milner PC, Payne JN, Lewis Pa, Jennison C. Inequalities in access to knee joint replacements for people in need. *Ann Rheum Dis*. 2004;63(11):1483–9.

163 Wolterbeek N, Garling EH, Mertens BJ, Nelissen RGHH, Valstar ER. Kinematics and early migration in single-radius mobile- and fixed-bearing total knee prostheses. *Clin Biomech*. 2012;27(4):398–402.

164 Roos EM. Effectiveness and practice variation of rehabilitation after joint replacement. *Curr Opin Rheumatol*. 2003;15(2):160–2.

165 Bini SA, Sidney S, Sorel M. Slowing demand for total joint arthroplasty in a population of 3.2 million. *J Arthroplasty*. 2011;26(6 Suppl):124–8.

166 Scott CEH, Bugler KE, Clement ND, Macdonald D, Howie CR, Biant LC. Patient expectations of arthroplasty of the hip and knee. *J Bone Joint Surg Br*. 2012;94(7):974–81.

167 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–15.

168 King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171–84.

169 Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage*. 2005;13(12):1076–83.

170 Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage*. 2007;15(3):273–80.

171 Nilsdotter AK, Aurell Y, Siösteen AK, Lohmander LS, Roos HP. Radiographic stage of osteoarthritis or sex of the patient does not predict one year outcome after total hip arthroplasty. *Ann Rheum Dis*. 2001;60(3):228–32.

172 Meding JB, Anderson AR, Faris PM, Keating EM, Ritter MA. Is the preoperative radiograph useful in predicting the outcome of a total hip replacement? *Clin Orthop*. 2000;(376):156–60.

173 Judge A, Javaid MK, Arden NK, Cushnaghan J, Reading I, Croft P, Dieppe PA, Cooper C. Clinical tool to identify patients who are most likely to achieve long-term improvement in physical function after total hip arthroplasty. *Arthritis Care Res*. 2012;64(6):881–9.

174 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. 1957;16(4):494–502.

175 Hobbs T. Twelve year report of The New Zealand Joint Registry.
URL `http://www.cdhb.govt.nz/njr/reports/A2D65CA3.pdf`

176 Gossec L, Paternotte S, Bingham CO, Clegg DO, Coste P, Conaghan PG, Davis AM, Giacovelli G, Gunther KP, Hawker G, Hochberg MC, Jordan JM, Katz JN, Kloppenburg M, Lanzarotti A, Lim K, Lohmander LS, Mahomed NN, Maillefert JF, Manno RL, March LM, Mazzuca SA, Pavelka K, Punzi L, Roos EM, Rovati LC, Shi H, Singh JA, Suarez-Almazor ME, Tajana-Messi E, Dougados M. OARSI/OMERACT initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. An OMERACT 10 Special Interest Group. *J Rheumatol*. 2011;38(8):1765–9.

177 de Beer J, Petruccelli D, Adili A, Piccirillo L, Wismer D, Winemaker M. Patient perspective survey of total hip vs total knee arthroplasty surgery. *J Arthroplasty*. 2012;27(6):865–9.e1–5.

178 Dobbs HS. Survivorship of total hip replacements. *J Bone Joint Surg Br*. 1980;62-B(2):168–73.

179 Dorey F, Amstutz HC. Survivorship analysis in the evaluation of joint replacement. *J Arthroplasty*. 1986;1(1):63–9.

180 Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389–430.

181 Yan Y, Moore R, Hoover D. Competing risk adjustment reduces overestimation of opportunistic infection rates in AIDS. *J Clin Epidemiol*. 2000;53(8):817–822.

182 Southern DA, Faris PD, Brant R, Galbraith PD, Norris CM, Knudtson ML, Ghali WA. Kaplan-Meier methods yielded misleading results in competing risk scenarios. *J Clin Epidemiol*. 2006;59(10):1110–4.

183 Fiocco M, Putter H, van de Velde CJH, van Houwelingen JC. Reduced rank proportional hazards model for competing risks: An application to a breast cancer trial. *J Stat Plan Inference*. 2006;136(5):1655–1668.

184 Fiocco M, Putter H, Van Houwelingen JC. Reduced rank proportional hazards model for competing risks. *Biostatistics*. 2005;6(3):465–78.

185 Bernoulli D. Essai d'une nouvelle analyse de la mortalite causee par la petite verole et des avantages de l'inoculation pour la prevenir. *Mém Math Phys Acad Roy Sci*. 1760;pp. 1–45.

186 Blower S, Bernoulli D. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. 1766. *Rev Med Virol*. 2004;14(5):275–88.

187 de Wreede LC, Fiocco M, Putter H. mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. *J Stat Softw*. 2011;38(7):1–30.

188 de Wreede LC, Fiocco M, Putter H. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed*. 2010;99(3):261–74.

189 Keurentjes JC, Kuipers RM, Wever DJ, Schreurs BW. High incidence of squeaking in THAs with alumina ceramic-on-ceramic bearings. *Clin Orthop Relat Res*. 2008; 466:1438–1443.

190 Harris WH, Sledge CB. Total hip and total knee replacement. *N Engl J Med*. 1990; 323(11):725–31.

191 Singh J, Sloan JA, Johanson NA. Challenges with health-related quality of life assessment in arthroplasty patients: problems and solutions. *J Am Acad Orthop Surg*. 2010;18(2):72–82.

192 Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77(4):371–83.

193 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–9.

194 Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower ex. *Arthritis Rheum*. 2001;45(4):384–391.

195 Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract*. 2000;6(1):39–49.

196 Cohen J. *Statistical power analysis for the behavioral sciences*, vol. 2. Erlbaum. 1988.

197 Ware J. *SF-36 health survey : manual and interpretation guide*. Boston: The Health Institute New England Medical Center. 1993.

198 Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. URL `http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp`

199 Shi HY, Mau LW, Chang JK, Wang JW, Chiu HC. Responsiveness of the Harris Hip Score and the SF-36: five years after total hip arthroplasty. *Qual Life Res*. 2009;18(8):1053–60.

200 Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis Cartilage*. 2007;15(9):981–1000.

201 Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage*. 2008;16(2):137–62.

202 Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage*. 2010;18(4):476–99.

203 Chesworth BM, Mahomed NN, Bourne RB, Davis AM. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol*. 2008;61(9):907–18.

204 Walther BA, Moore JL. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*. 2005;28(6):815–829.

205 Arden NK, Kiran A, Judge A, Biant LC, Javaid MK, Murray DW, Carr AJ, Cooper C, Field RE. What is a good patient reported outcome after total hip replacement? *Osteoarthritis Cartilage*. 2011;19(2):155–62.

206 Tubach F, Wells GA, Ravaud P, Dougados M. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. *J Rheumatol*. 2005;32(10):2025–9.

207 Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D, Dougados M. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis*. 2005;64(1):34–7.

208 Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78:185–190.

209 Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br*. 1998;80:63–69.

210 Judge A, Arden NK, Kiran A, Price A, Javaid MK, Beard D, Murray D, Field RE. Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery. *J Bone Joint Surg Br*. 2012;94(3):412–8.

211 Gosens T, Hoefnagels NHM, de Vet RCW, Dhert WJA, van Langelaan EJ, Bulstra SK, Geesink RGT. The "Oxford Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthop*. 2005;76(2):204–11.

212 Haverkamp D, Breugem SJM, Sierevelt IN, Blankevoort L, van Dijk CN. Translation and validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty. *Acta Orthop*. 2005;76(3):347–52.

213 Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149–58.

214 Judge A, Arden NK, Price A, Glyn-Jones S, Beard D, Carr AJ, Dawson J, Fitzpatrick R, Field RE. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? *J Bone Joint Surg Br*. 2011;93(12):1660–4.

215 Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis*. 2007;66 Suppl 3(suppl_3):iii40–1.

216 Clement ND, Macdonald D, Simpson AHRW. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. 2013;22(8):1933–9.

217 Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop*. 2014;85(1):39–48.

218 Browne JP, van der Meulen JH, Lewsey JD, Lamping DL, Black N. Mathematical coupling may account for the association between baseline severity and minimally important difference values. *J Clin Epidemiol*. 2010;63(8):865–74.

219 Dougados M. It's good to feel better but it's better to feel good. *J Rheumatol*. 2005; 32(1):1–2.

220 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2(14):i–iv, 1–74.

221 Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. *J Bone Joint Surg Br*. 2011;93(7):867–75.

222 Kvien TK, Mowinckel P, Heiberg T, Dammann KL, Dale O, Aanerud GJ, Alme TN, Uhlig T. Performance of health status measures with a pen based personal digital assistant. *Ann Rheum Dis*. 2005;64(10):1480–4.

223 Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S, Davis T. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. *J Clin Epidemiol*. 2011;64(2):182–90.

224 Manfreda K, Bosnjak M, Berzelak J, Haas I, Vehovar V. Web surveys versus other survey modes: A meta-analysis comparing response rates. *Int J Market Res*. 2008; 50(1):79–104.

225 Rolfson O, Salomonsson R, Dahlberg LE, Garellick G. Internet-based follow-up questionnaire for measuring patient-reported outcome after total hip replacement surgery-reliability and response rate. *Value Health*;14(2):316–21.

226 Ryan JM, Corry JR, Attewell R, Smithson MJ. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Qual Life Res*. 2002;11(1):19–26.

227 Evangelou E, Kerkhof HJ, Styrkarsdottir U, Ntzani EE, Bos SD, Esko T, Evans DS, Metrustry S, Panoutsopoulou K, Ramos YFM, Thorleifsson G, Tsilidis KK, Arden N, Aslam N, Bellamy N, Birrell F, Blanco FJ, Carr A, Chapman K, Day-Williams AG, Deloukas P, Doherty M, Engström G, Helgadottir HT, Hofman A, Ingvarsson T, Jonsson H, Keis A, Keurentjes JC, Kloppenburg M, Lind PA, McCaskie A, Martin NG, Milani L, Montgomery GW, Nelissen RGHH, Nevitt MC, Nilsson PM, Ollier WE, Parimi N, Rai A, Ralston SH, Reed MR, Riancho JA, Rivadeneira F, Rodriguez-Fontenla C, Southam L, Thorsteinsdottir U, Tsezou A, Wallis GA, Wilkinson JM, Gonzalez A, Lane NE, Lohmander LS, Loughlin J, Metspalu A, Uitterlinden AG, Jonsdottir I, Stefansson K, Slagboom PE, Zeggini E, Meulenbelt I, Ioannidis JP, Spector TD, van Meurs JBJ, Valdes AM. A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis*. 2013;0:1–7.

228 Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika*. 1991;78(3):691–692.

229 Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 2000;19(8):1059–79.

230 Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441–7.

231 Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. Springer. 2009.

232 Deutskens E, de Ruyter K, Wetzels M, Oosterveld P. Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Mark Lett*. 2004;15(1):21–36.

233 Brandtzaeg PB, Heim J, Karahasanović A. Understanding the new digital divide: A typology of Internet users in Europe. *Int J Hum Comput Stud*. 2011;69(3):123–138.

234 Lee SJ, Kavanaugh A, Lenert L. Electronic and computer-generated patient questionnaires in standard care. *Best Pract Res Clin Rheumatol*. 2007;21(4):637–47.

235 Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RGHH. Good diagnostic performance of early migration as a predictor of late aseptic loosening of acetabular cups: results from ten years of follow-up with Roentgen stereophotogrammetric analysis (RSA). *J Bone Joint Surg Am*. 2012;94(10):874–80.

236 Mancuso CA, Graziano S, Briskie LM, Peterson MGE, Pellicci PM, Salvati EA, Sculco TP. Randomized trials to modify patients' preoperative expectations of hip and knee arthroplasties. *Clin Orthop*. 2008;466(2):424–31.

237 Ranstam J, Kärrholm J, Pulkkinen P, Mäkelä K, Espehaug B, Pedersen AB, Mehnert F, Furnes O. Statistical analysis of arthroplasty data. II. Guidelines. *Acta Orthop*. 2011; 82(3):258–67.

238 Werner F, Werner K, Oberzaucher J. *Tablets for Seniors: An Evaluation of a Current Model (iPad)*. Advanced Technologies and Societal Change. Berlin, Heidelberg: Springer Berlin Heidelberg. 2012.

Dutch Summary

Achtergrond

In dit proefschrift werden voorspellers van de klinische uitkomst na een totale heup- en knieprothese onderzocht. De klinische uitkomst kan op verschillende wijzen en van verschillende perspectieven benaderd worden. Grofweg kan de volgende driedeling worden gehanteerd: het medisch-technische perspectief, het patiëntperspectief en het maatschappelijke perspectief. In het eerste deel (hoofdstuk 2) lag de nadruk op het medisch-technische aspect. Het patiëntperspectief kwam aan bod in deel twee (hoofdstuk 3 en 4). Tenslotte werden in het derde deel (hoofdstuk 5 – 9) methodologische aspecten behandeld, die ten grondslag liggen aan orthopaedisch onderzoek. Het maatschappelijke aspect, welk beoordeeld kan worden binnen het raamwerk van de International Classification of Functioning, Disability and Health, wordt in dit proefschrift grotendeels buiten beschouwing gelaten.

Een belangrijke medisch-technische uitkomstmaat is de kans dat een patiënt binnen 10 jaar na plaatsing van de prothese een heringreep moet ondergaan wegens het falen van het implantaat. Deze kans vormt de basis waarop de verschillende totale heup- en knieprotheses onderling vergeleken worden. Over het algemeen geldt dat een goed medisch-technisch resultaat een randvoorwaarde is voor een goede uitkomst, ook vanuit het perspectief van de patiënt. Toch is 10–30% van de patiënten na een totale heup- of knieprothese ontevreden met de uitkomst, ondanks een medisch-technisch goed resultaat.

Om meer inzicht te krijgen in het perspectief van de patiënt, zoals de toename in kwaliteit van leven en de tevredenheid met het operatieresultaat, wordt er gebruik gemaakt

van vragenlijsten. Deze vragenlijsten, ook wel Patient Reported Outcome Measures (PROMs) genoemd, worden door de patiënt zelf ingevuld. Dergelijke meetinstrumenten kunnen in twee categorieën ingedeeld worden: generieke en ziektebeeld-specifieke vragenlijsten. In dit proefschrift lag de nadruk op generieke vragenlijsten. Om onderzoek te kunnen doen naar de toename in kwaliteit van leven na een heup- en knieprothese, hebben we de Paprika Studie opgezet: *Pa*tiënten *Pr*ospectief gevolgd *i*n het kader van een totale *k*nie- of heup*a*rthoplastiek (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189; Netherlands Trial Register: NTR2190). Patiënten die eerder deel hebben genomen aan de Trigger Studie of de TOMaat Studie, werden benaderd om deel te nemen aan een vragenlijstonderzoek en aan genetisch onderzoek, om een basis te vormen voor associatiestudies tussen genotype en loslating van heup- en knieprothesen.

Samenvatting

In hoofdstuk 1 (p. 1) werd de epidemiologie van heup- en knieprothesen in Nederland besproken. Daarnaast werd een kort overzicht gegeven van de opzet van de Paprika Studie en een overzicht gegeven van de inhoud van dit proefschrift.

In hoofdstuk 2 (p. 9) hebben we een systematische literatuurstudie verricht naar de kans op een heringreep binnen 10 jaar na implantatie voor elk type primaire heupimplantaat. We vergeleken de uitkomsten van elk implantaat met de NICE benchmarks.[35] We vonden 10-jaars resultaten van 34 verschillende typen acetabulaire cups en van 32 verschillende femorale stelen. De gepubliceerde resultaten van 8 acetabulaire cups en 15 femorale stelen waren significant beter dan de NICE benchmarks; Van 16 acetabulaire cups en 6 femorale stelen waren de gepubliceerde resultaten significant slechter dan de NICE benchmarks. De methodologische kwaliteit van de meeste onderzoeken was laag. Dit betekent dat het risico op gebiasde resultaten hoog is. De conclusies van de beoordeelde artikelen wordt hierdoor twijfelachtig.

In hoofdstuk 3 (p. 37) hebben we onderzocht of de socio-economische positie van de patiënt van invloed is op de verbetering in kwaliteit van leven en de postoperatieve patiënttevredenheid na een totale heup- of knieprothese. We hebben de deelnemers

van de Paprika Studie in drie groepen onderverdeeld, op basis van de hoogst voltooide opleiding, wat een goede indicator is van de socio-economische positie. We vonden een vergelijkbare toename in kwaliteit van leven en een vergelijkbare patiënttevredenheid voor elke groep na een totale heupprothese. Na een totale knieprothese vonden we kleine verschillen in toename in kwaliteit van leven tussen de groepen, zonder klinische relevantie. De patiënttevredenheid was vergelijkbaar voor elke groep. De socio-economische positie van de patiënt is geen goede voorspeller voor de toename in kwaliteit van leven en patiënttevredenheid na een heup- en knieprothese.

In hoofdstuk 4 (p. 55) hebben we onderzocht of de pre-operatieve radiologische gradatie van artrose gerelateerd is aan de verbetering in kwaliteit van leven na een totale heup- of knieprothese. De pre-operatieve radiologische artrose van deelnemers aan de Paprika Studie werd beoordeeld volgens Kellgren en Lawrence[174] en ingedeeld in twee groepen: patiënten met ernstige artrose (Kellgren en Lawrence graad 3 en 4) en patiënten met milde artrose (Kellgren en Lawrence graad 0, 1 of 2). Patiënten met een ernstige artrose verbeterden niet alleen meer, maar hadden ook een hogere kans op een relevante verbetering van het fysieke functioneren na zowel een totale heup- als knieprothese. De patiënttevredenheid was ook hoger na een totale knieprothese bij patiënten met ernstige radiologische artrose. De pre-operatieve radiologische artrose kan dus een nuttige variabele zijn om de toename in fysiek functioneren en postoperatieve tevredenheid te voorspellen.

In hoofdstuk 5 (p. 71) hebben we onderzocht hoeveel bias er geïntroduceerd wordt in de schatting van de kans op een heroperatie na een heupprothese, wanneer een cruciale aanname van de Kaplan-Meier analyse wordt geschonden. In de Kaplan-Meier analyse neemt men aan dat de tijd tot het event onafhankelijk is van het mechanisme van censurering. Dit is een plausibele aanname in de oorspronkelijke toepassing van de Kaplan-Meier analyse, namelijk het schatten van de cumulatieve kans om nog in leven te zijn op elk willekeurig moment, in een populatie waarin niet iedereen overleden is bij het einde van de follow-up van de studie (oftewel rechts-censurering). Bij andere eindpunten, zoals de kans op een heroperatie, kunnen er competing events

optreden: gebeurtenissen die ervoor zorgen dat het event niet meer plaats kan vinden. In het geval van heupprothesen is het overlijden van de patiënt een treffend voorbeeld van een competing event: de heupprothese van een overleden patiënt zal niet meer gereviseerd worden. Wanneer er competing events optreden, wordt de aanname van onafhankelijkheid van de tijd tot event en het mechanisme van censurering geschonden. De Kaplan-Meier analyse leidt in aanwezigheid van competing events áltijd tot een overschatting van de cumulatieve incidentie van het event.

In hoofdstuk 6 (p. 81) hebben we een systematische literatuurstudie verricht naar Minimal Clinically Important Differences (MCIDs) in kwaliteit van leven scores (gemeten middels Short-Form 36) na een totale heup- en knieprothese. Door middel van meta-analyse hebben we getracht de precisie van MCIDs te vergroten. Helaas bleek dit niet mogelijk: we vonden 3 schattingen van 3 verschillende populaties (primaire totale heupprothese, primaire totale knieprothese, revisie totale heupprothese). De gevonden MCID schattingen waren niet gevalideerd met behulp van externe criteria, gerelateerd aan bijvoorbeeld patiënttevredenheid. Daarnaast hadden de gevonden MCID schattingen een geringe precisie, wat herleidbaar was uit de wijde betrouwbaarheidsintervallen. Tot op heden zijn dit echter de best bekende schattingen. We adviseren enige terughoudendheid in het gebruik van deze MCIDs als absolute drempelwaarden in de beoordeling of een patiënt al dan niet een minimaal klinisch relevante toename in kwaliteit van leven heeft ondervonden na een heup- of knieprothese.

In hoofdstuk 7 (p. 93) hebben we op een innovatieve wijze Clinically Important Differences (CID) in kwaliteit van leven (gemeten middels SF36) geschat na een totale heup- en knieprothese, gebruik makend van data van de Paprika Studie. CIDs zijn drempelwaarden die een substantiële verbetering aangeven na een interventie. Voor de kliniek zijn deze drempelwaarden relevanter dan MCIDs, aangezien men ook een substantiële verbetering verwacht van een totale heup- of knieprothese. CID schattingen van de subschalen Physical Functioning, Role Physical, Bodily Pain en Social Functioning werden gevalideerd door de volgende validatievraag: “Zou u

deze ingreep opnieuw willen ondergaan, nu u weet hoe de resultaten voor u zijn?". We adviseren enige terughoudendheid in het gebruik van de CIDs van de overige subschalen als absolute drempelwaarden, omdat deze niet extern gevalideerd zijn.

In hoofdstuk 8 (p. 105) hebben we Patient Acceptable Symptom States (PASS) van de Oxford Hip Score (OHS) en de Oxford Knee Score (OKS) geschat na een totale heup- en knieprothese, gebruik makend van data van de Paprika Studie. PASS zijn drempelwaarden, waarboven een patiënt een acceptabele uitkomst heeft bereikt. Receiver Operating Characteristic curves identificeerden een optimale balans tussen sensitiviteit en specificiteit bij een PASS drempel van 42 punten voor de OHS na een totale heupprothese en 37 punten voor de OKS na TKR. Heupprothese patiënten met een OHS $\geq$ 42 punten en knieprothese patiënten met een OKS $\geq$ 37 punten hadden een hogere tevredenheidsscore en gaven vaker aan dat ze bereid waren de operatie opnieuw te ondergaan, gegeven hun uitkomst. De PASS drempelwaarden verschilden echter aanzienlijk tussen relevante subgroepen. In vergelijking met gepubliceerde PASS drempelwaarden 6 maanden na de ingreep, lijken de PASS drempelwaarden na gemiddeld 3 jaar na de operatie hoger. We adviseren deze PASS drempelwaarden eerst te valideren in een externe populatie, alvorens ze te gebruiken in de praktijk bij het bepalen of een patiënt een aanvaardbare uitkomst na een totale heup- of knieprothese heeft bereikt.

In hoofdstuk 9 (p. 119) hebben we de voorkeur voor het gebruik van een type vragenlijst geïnventariseerd van deelnemers aan de Paprika Studie: gebruik van papieren vragenlijsten of vragenlijsten elektronisch invullen en per email terugsturen. Een ruime meerderheid ($>$80%) van de patiënten gaf aan de papieren vragenlijst te verkiezen boven een elektronische variant. Patiënten die de voorkeur gaven aan een elektronische vragenlijst, waren gemiddeld jonger en hoger opgeleid. Onderzoek doen naar kwaliteit van leven na heup- of knieprothesen met uitsluitend elektronische vragenlijsten kan leiden tot selectie bias.

Tenslotte werden de bevindingen van dit proefschrift samengevat in hoofdstuk 10 en werden er suggesties gedaan voor vervolgonderzoek.

Acknowledgements

I would like to thank all patients who participated in the Paprika Study and all patients who participated in the studies which were summarised in the systematic reviews in this thesis. Without their kind effort and perseverance none of these studies would have been possible.
I am deeply grateful to the Dutch Arthritis Foundation for their grant, which supported my position as a PhD-student at the Leiden University Medical Center.

Additionally, I would like to thank prof. dr. Rob G. Nelissen for stimulating me to develop my academic potential and for the thought-provoking and entertaining discussions on a broad range of subjects, such as medicine, science in general, politics and arts; dr. Marta Fiocco for sharing my enthusiasm for innovative methodology and for the privilege of her private tutoring in biostatistics; the staff of the department of clinical epidemiology at the Leiden University Medical Center for providing a sound epidemiological base; dr. B. Wim Schreurs for introducing me to the world of scientific research; prof. dr. Thea P. M. Vliet Vlieland for her assistance and support in improving my organisational skills and Gerda Kuipers for her support in study logistics.

Finally, I would like to thank my colleagues, friends and family for their personal support.

Curriculum Vitae

J. Christiaan Keurentjes was born on the 14th of August 1982 in Satu Mare, Romania and raised in Doetinchem, The Netherlands. He graduated from secondary school in 2001 at the Rietveld Lyceum in Doetinchem. In 2001 he started his medical training at the Radboud University Medical Center in Nijmegen, The Netherlands. During his clinical rotations, he set up research projects under the supervision of dr. B. Wim Schreurs, resulting in a number of publications.

In 2009, Christiaan started working as a PhD student at the department of orthopaedic surgery at the Leiden University Medical Center, under the supervision of prof. dr. Rob G. Nelissen. He participated in a training programme at the department of clinical epidemiology at the Leiden University Medical Center, which will result in a PhD in clinical epidemiology. He is an invited speaker for several international conferences on health-related quality of life and a peer reviewer for a number of leading journals in orthopaedic surgery, epidemiology and general medicine.

After having worked as a resident in emergency medicine (2008: Winterswijk, The Netherlands), orthopaedic surgery (2009: Hilversum, The Netherlands; 2012: Gouda, The Netherlands) and pathology (2013: Nijmegen, The Netherlands), Christiaan decided to pursue a career in clinical epidemiology. He founded an Epidemiological Consultancy, aiming to improve both health care and epidemiology through innovative clinical research.