Minimal clinically important differences in health-related quality of life after total hip or knee replacement

A SYSTEMATIC REVIEW

J. C. Keurentjes, F. R. Van Tol, M. Fiocco, J. W. Schoones, R. G. Nelissen

From Leiden University Medical Center, Leiden, The Netherlands

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.
undergoing an intervention. 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. 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. 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. 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, 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. The within-person global change rating is often used as an anchor, which is measured using Likert scales, ranging from five to 15 options. 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’. 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. 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. Generally accepted individual effect size standards are equal to the group effect size standards, as defined by Cohen. Therefore, the MCID is calculated by multiplying the standard deviation of patients at baseline by 0.5.

Recently Quintana et al and Escobar et al 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.

Materials and 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. 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 MCID[tw] OR “Minimal clinically important changes” OR “Minimal clinically important change” OR “clinically important changes” OR “Minimal clinical important changes” OR “Minimal clinical important difference” OR “Minimal clinical important difference” OR “clinical important differences” OR “clinical important differences” OR “clinical important change” OR “minimal detectable change” OR “minimal detectable changes” OR “minimal detectable change” 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
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, 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., 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. 1). 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.
Study characteristics. An overview of the study characteristics of all included studies is presented in Table I. Quintana et al\textsuperscript{10} describe the MCID in SF-36 after primary THR at follow-up periods of six months and two years; Escobar et al\textsuperscript{11} describe the MCID in SF-36 after primary TKR at follow-up periods of six months and two years; and Shi et al\textsuperscript{16} describe the MCID in SF-36 after revision THR at a follow-up period of six months (Table I). 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\textsuperscript{10}; 76 after six months and 65 after two years for Escobar et al\textsuperscript{11}; and 67 after six months for Shi et al\textsuperscript{16} (Table I). The indication for joint replacement was osteoarthritis in all patients of Quintana et al\textsuperscript{10} and Escobar et al\textsuperscript{11}, while Shi et al\textsuperscript{16} offered no statement of the indication for joint replacement (Table I). 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 II. Two studies used anchor-based approaches to estimate the MCID\textsuperscript{10,11} while the other used a distribution-based approach\textsuperscript{16}. No study performed any form of additional validation. The study populations of Quintana et al\textsuperscript{10} and Escobar et al\textsuperscript{11} are truly representative of THR and TKR patients in general, while Shi et al\textsuperscript{16} 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 III and Figures 2 and 3. The MCIDs are estimated with 95% confidence intervals for each of the SF-36 domains in primary TKR and primary and revision THR at six months (Fig. 2, Table III)\textsuperscript{10,11,16} and for primary TKR and THR at two years post-operatively (Fig. 3, Table III)\textsuperscript{10,11}.

### Table I. Study characteristics of the three included studies

| Authors          | Intervention | Follow-up | Setting       | Positive/ negative MCID estimated | Sample size of study | Sample size for MCID estimation | Osteoarthritis (%) | Male gender (%) | Mean age (yrs) | Lost to follow-up (%) |
|------------------|--------------|-----------|---------------|----------------------------------|----------------------|-------------------------------|-------------------|----------------|----------------|----------------------|
| Escobar et al\textsuperscript{11} | Primary TKR  | 6 months  | Multi-centre  | Positive MCID only | 423                  | 76                            | 100               | 25.0           | 71.6          | 22.39                |
|                  | Primary TKR  | 2 years   | Multi-centre  | Positive MCID only | 364                  | 65                            | 100               | 25.0           | 71.6          | 33.21                |
| Quintana et al\textsuperscript{10} | Primary THR  | 6 months  | Multi-centre  | Positive MCID only | 485                  | 43                            | 100               | 49.3           | 69.4          | 21.86                |
|                  | Primary THR  | 2 years   | Multi-centre  | Positive MCID only | 310                  | 33                            | 100               | 49.3           | 69.4          | 36.08                |
| Shi et al\textsuperscript{16} | Revision THR | 6 months  | Multi-centre  | Positive MCID only | 67                   | 67                            | n/a               | 56.7           | 70.2          | 16.25                |

### Table II. Risk of bias within the three included studies

| Authors          | Intervention\* | Follow-up | MCID\† methodology | Additional validation | Representativeness of study population | Adequacy of follow-up |
|------------------|-----------------|-----------|---------------------|-----------------------|---------------------------------------|-----------------------|
| Escobar et al\textsuperscript{11} | Primary TKR     | 6 months  | Anchor-based        | No                    | Truly representative                    | More than 5% lost, results possibly biased |
|                  | Primary TKR     | 2 years   | Anchor-based        | No                    | Truly representative                    | More than 5% lost, results possibly biased |
| Quintana et al\textsuperscript{10} | Primary THR    | 6 months  | Anchor-based        | No                    | Truly representative                    | More than 5% lost, results possibly biased |
|                  | Primary THR    | 2 years   | Anchor-based        | No                    | Truly representative                    | More than 5% lost, results possibly biased |
| Shi et al\textsuperscript{16} | Revision THR   | 6 months  | Distribution-based  | No                    | Description of cohort incomplete: no statement on indication for primary intervention | More than 5% lost, results possibly biased |

\* TKR, total knee replacement; THR, total hip replacement

\† MCID, minimal clinically important difference
Discussion

We have found one study describing MCIDs in SF-36 after primary THR, one after primary TKR and one after revision THR; 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.

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 Figures 2 and 3.

The findings of this systematic review underline the need to identify MCIDs for each specific population. As can be seen from Figures 2 and 3, MCIDs differ both between SF-36 subscales and patient populations. The use of a ‘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. 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. 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. and Escobar et al. 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.

A strongly recommended method of determining MCIDs is by triangulation of multiple approaches. None of the included studies has applied any form of additional validation, such as secondary anchor questions; all used a.

Table III. 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)

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

References
1. Harris WH, Sledge CB. Total hip and total knee replacement (1). N Engl J Med 1990;323:725–731.
2. 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;18:72–82.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–415.
4. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171–184.
5. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371–383.
6. 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;61:102–109.
7. 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 extremities. Arthritis Rheum 2001;45:384–391.
8. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000;6:39–49.
9. Cohn J. Statistical power analysis for the behavioral sciences. New York: Routledge, 1988.
10. Quintana JM, Escobar A, Bilbao A, et al. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. Osteoarthritis Cartilage 2005;13:1076–1083.
11. Escobar A, Quintana JM, Bilbao A, et al. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. Osteoarthritis Cartilage 2007;15:273–280.
12. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
13. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute New England Medical Center, 1993.
14. Wells GA, Shea B, O’Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (date last accessed 2 May 2012).
15. Pijls BG, Dekkers OM, Middeldorp S, et al. AQUIA: assessment of quality in lower limb arthroplasty: an expert Delphi consensus for total knee and total hip arthroplasty. BMC Musculoskelet Disord 2010;11:261.
16. Shi HY, Chang JK, Wong CY, et al. Responsiveness and minimal important differences after revision total hip arthroplasty. BMC Musculoskelet Disord 2010;11:261.
17. Zhang W, Moskowitz RW, Nuki G, et al. 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;15:981–1000.
18. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16:137–162.
19. Zhang W, Nuki G, Moskowitz RW, et al. 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;18:476–499.
20. Chesworth BM, Mahomed NN, Bourne RB, Davis AM; OJRR Study Group. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. J Clin Epidemiol 2008;61:907–918.
21. 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;28:815–829.