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

Keurentjes, J.C.

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.

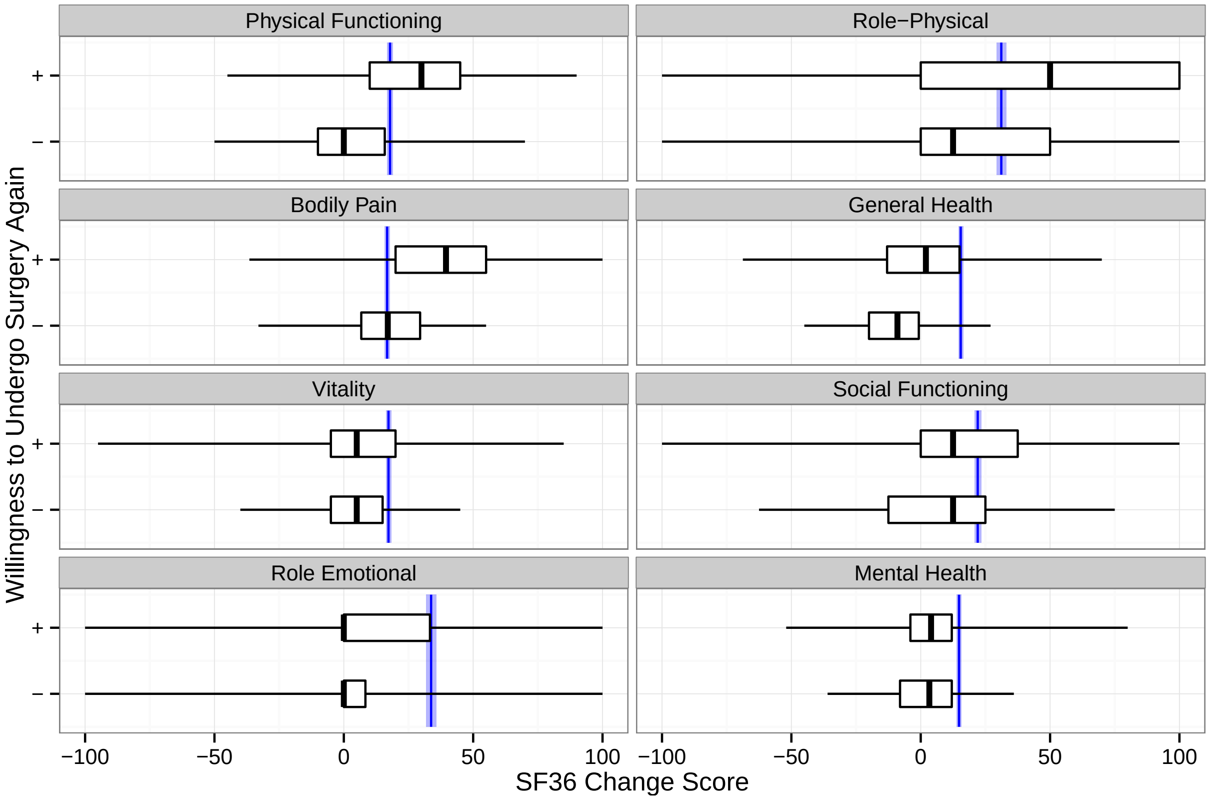

Figure 7.1: **Improvement in HRQoL after THR per Validation Question.** The vertical blue lines indicate the CID of each sub-scale with its confidence interval shown in purple; the boxplots indicate the median, IQR and range of patients, who answered the Validation Question positively and negatively.

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,

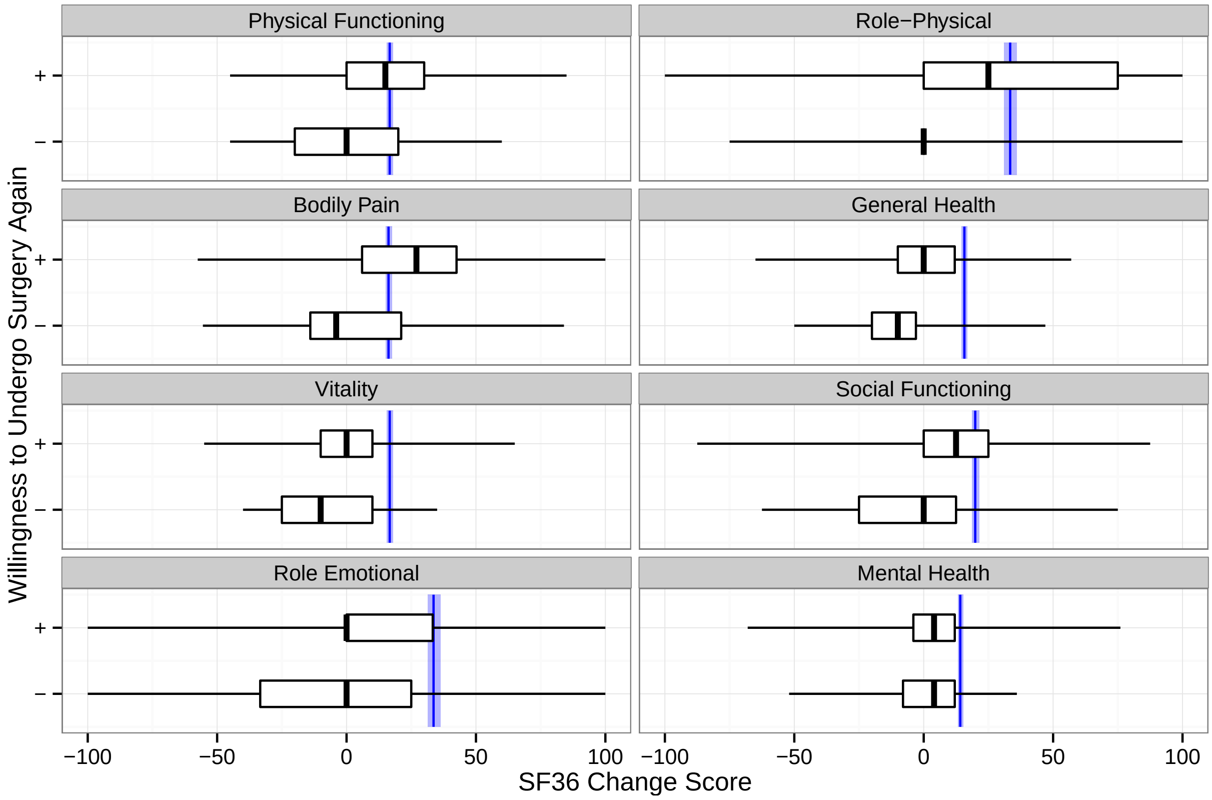

Figure 7.2: **Improvement in HRQoL after TKR per Validation Question.** The vertical blue lines indicate the CID of each sub-scale with its confidence interval shown in purple; the boxplots indicate the median, IQR and range of patients, who answered the Validation Question positively and negatively.

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.