Short-term patient-reported outcomes following total hip replacement: Is the success picture overrated?

Marek Kamil Gojło a, Robert Lundqvist b, Przemysław T. Paradowski c, d, e, *

a Department of Orthopaedics and Traumatology, Ministry of Interior and Administration Hospital, Warmia and Masurian Oncology Center, Wojska Polskiego 37, PL-10-228, Olsztyn, Poland
b Research and Innovation Unit, Norrbotten County Council, SE-971 89, Luleå, Sweden
c Department of Surgical and Perioperative Sciences, Division of Orthopedics, Sunderby Research Unit, Umeå University, Sunderby Central Hospital of Norrbotten, SE-971 80, Luleå, Sweden
d Faculty of Health Sciences, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University in Toruń, Jagellonka 13-15, PL-85-067, Bydgoszcz, Poland
e Clinical Epidemiology Unit, Orthopedics, Department of Clinical Sciences Lund, Lund University, SE-221 85, Lund, Sweden

ARTICLE INFO

Keywords: Hip Osteoarthritis Arthroplasty Replacement HOOS Outcome

ABSTRACT

Objectives: Our objective was report short-term results of total hip replacement (THR) and to identify patients with functional recovery and with treatment failure. We also aimed to investigate whether there are any potential predicting factors for functional recovery or treatment failure.

Design: Prospective cohort study. Clinical examination was performed and data were collected from patients before THR and at three follow-up assessments within subsequent year. The primary endpoint was the change between assessments in the average score on four subscales of the Hip disability and Osteoarthritis Outcome Score (HOOS) covering pain, symptoms, activity of daily living, and quality of life. Secondary endpoints included results on all five HOOS subscales, and the SF–36. Functional Recovery and Treatment Failure were defined basing on the published reference population values.

Results: We assessed 179 patients (98 women and 81 men, mean age at THR 67 years, range 31–90 years). The mean HOOS scores continued to improve up to 12 months after THR. Functional Recovery was identified in 32% while Treatment Failure in 16% of the patients. We found an association between high (>30.3) preoperative SF–36 Physical Component Summary scores and Functional Recovery, as well as low preoperative SF–36 Physical (<30.3) and Mental (<35.9) Component Summary scores and Treatment Failure.

Conclusions: Knowing that only one third of subjects undergoing THR achieved Functional Recovery and one sixth had Treatment Failure, gives us a better perspective to discuss feasibility of expectations and, consequently, to prevent patient dissatisfaction following THR.

1. Introduction

Total hip replacement (THR) is one of the most common and most successful surgical operations in medicine [1]. THR is undertaken to relieve pain and restore function to the hip joint with end-stage osteoarthritis (OA) [2]. Traditionally, the outcome of THR has been quantified by orthopaedic surgeons as morbidity and mortality rates, alternatively as implant survival rate indices [3,4]. The 10-year survival of the THR implant was reported to be as high as 95% and 15-year survival over 85% [5]. The implant survival index, however, says nothing about postoperative patient satisfaction, pain reduction, hip function or quality of life. It has been estimated that between 6 and 15% of patients undergoing THR have persistent pain and functional limitation [6] and 6–7% are not satisfied with the result of the operation after one year [7,8]). Since patients are generally less satisfied with their outcomes than surgeons [9], assessments made from the patient’s perspective with patient-reported outcome measures (PROMs) are advocated [10]. PROMs are now commonly used in arthroplasty registers [11]. Most of the studies involving THR patients THR, even those based on registers, report patients’ outcomes in a longer time perspective [12]. It has been found, however, that most of clinically relevant improvement following primary joint replacement occurs within the first six months.
postoperatively [15,14] and that further progress on perceived physical functioning (but not on daily activities) can possibly be expected even beyond this period [14–17]. One-year follow-up scores are supposed to be two to three times higher than the baseline scores assessed before THR in different PROMs [18,19]. Although the population reference values for both Oxford and Harris Hip Scores (OHS, HSS) [20] and Hip disability and Osteoarthritis Outcome Score (HOOS) [21] have recently been published, they have not yet been referred to patients’ results after THR.

There is no consistency about what factors affect specific outcomes following THR [12].

It is known that worse preoperative function is associated with larger postoperative improvement [22–24]. However, patients with low baseline scores do not reach the level of postoperative functioning that subjects with better preoperative function do [25]. The results concerning the role of potential predictors such as comorbidity, pain and preoperative health-related quality of life, educational level, patient expectations and mental well-being are conflicting [12].

To our knowledge, no study has characterized results of THR based on reference population data for the HOOS. Thus, our primary purpose was to report short-term THR results and identify patients with good outcome corresponding to functional recovery, as well as patients with unsatisfactory results that can be consistent with treatment failure. The secondary purpose was to investigate whether there are any potential predicting factors for functional recovery or treatment failure after THR.

2. Methods

2.1. Study design

This is a prospective cohort trial in patients participating in the Study Evaluating the Efficacy of Joint Replacement (SEVERE). The study is registered on ClinicalTrials.gov (NCT04691466).

2.2. Patients

Patients who had undergone THR at the Department of Orthopaedics, Ministry of the Interior and Administration Hospital in Olsztyn, Poland, between April 2013 and April 2016 were included in the study. All the patients had either primary or secondary hip OA, according to the American College of Rheumatology criteria [26], with indication for THR.

Exclusion criteria were: 1) inability to understand the Polish language, 2) presence of neuromuscular disease and cognitive impairment, and 3) having a prosthesis in another joint of the ipsilateral or contralateral lower limb placed within 6 months before THR surgery, 4) having rheumatoid arthritis, and 5) symptoms in several joints (hip, knee or ankle) with expected total joint arthroplasty within 1 year.

Surgical procedure and strategy of the postoperative management are presented in Additional File 2.

2.3. Outcome measures

Condition-specific scores were assessed using the Hip disability and Osteoarthritis Outcome Score (HOOS). The HOOS is a 40-item questionnaire which consists of five subscales: Pain, other Symptoms, Function in Daily Living (or Activity in Daily Living, ADL Function), Sports and Recreation Function and hip related Quality of Life (QOL). A normalized score from 0 (indicating extreme problems) to 100 (indicating no problems at all) is calculated separately for each subscale.

The user’s guide can be downloaded from www.koos.nu [27]. The Polish version of the HOOS is considered to be a valid and reliable measure in the assessment of patient-relevant outcome in subjects undergoing THR [28].

General health quality of life was assessed by the 36-item Short Form Health Survey (SF-36) that has already been validated in Polish [29]. Scores from 0 (worst possible health status) to 100 (best possible health status) were independently generated for Physical Component Summary and Mental Component Summary.

SF-36 outcomes were calculated with Scoring Software v. 4.5 delivered by the copyrights holder (Optum Insight, Eden Prairie, MN, USA, license number QM018125). The Physical and Mental Summary Scores were calculated according to the oblique rotation analysis. The oblique model was more reasonable than orthogonal rotation model [30] since it permits the assumption that mental and physical health interact with each other [31].

The intensity of hip pain was measured with the visual analog scale (VAS). The 100-mm VAS is a unidimensional scale and it is considered valid and reliable [32].

2.4. Data collection

Data were collected prospectively at four time points: before THR (at baseline, assessment A), and at routine follow-up 6–8 weeks after THR (assessment B), 6 months after THR (assessment C) and, finally, 12 months after THR (assessment D).

The preoperative (baseline) and follow-up assessments were done in the clinic. During assessment A, the participants were asked to fill out the Polish version of HOOS, the SF-36, the VAS for pain, and give data on sex, age, number of joints affected by arthritis, level of education, and occupational status (Additional File 3). At the follow-up assessments, the participants were asked to fill out the HOOS questionnaire. All self-reported questionnaires, demographics and relevant information were processed by one orthopedic surgeon (MKG).

2.5. Primary outcome

2.5.1. HOOS4 score change

The primary outcome was the change between the HOOS4 assessments, with the scores ranging from 0 (worst) to 100 (best). The HOOS4 is defined as an average score of the four HOOS subscale scores: HOOS Pain, HOOS Symptoms, HOOS ADL, and HOOS QOL. The fifth subscale, Sports and Recreation Function is excluded to avoid a floor effect, since hardly any elderly THR patients undertake sports activities [33].

2.6. Secondary outcomes

Secondary outcomes included results of all five HOOS subscales, the scores of the SF-36 Physical Component Summary, and Mental Component Summary and pain intensity in VAS.

For the purpose of the analysis, two concepts: ‘Functional Recovery’ and ‘Treatment Failure’ have been created.

2.7. Functional Recovery and Treatment Failure

We have specified two categories of the HOOS outcome for the study group: ‘Functional Recovery’ or ‘Treatment Failure’. Based on the published Swedish population reference, the Functional Recovery was defined as the score above the lower threshold for the 95% CI in every single HOOS subscale in respective age groups (18–34, 35–54, 55–74 and over 75 years), separately for men and women [21].

Treatment Failure was defined as a HOOS QOL score <44 which is a prespecified cutoff value consistent with a report of more than a moderately decreased hip-related quality of life.

2.8. Statistical analysis

Descriptive statistics were used to describe sociodemographic and clinical characteristics preoperatively, at baseline, and clinical characteristics after treatment, at follow-up. Continuous data were described using means with standard deviations (SD) and medians. The Shapiro-Wilk test was used to determine if the data were normally distributed.
If the normality assumptions were met, parametric tests were used: independent or paired Student's t-test for quantitative variables and McNemar or Chi² test for qualitative variables. When the normality assumptions were not satisfied, Mann-Whitney U test for continuous variable and the Fisher's exact or McNemar test for binary variables were used.

For the analysis of HOOS₄, a linear mixed model (LMM) has been chosen. The reason is that with repeated values for the HOOS₄ variable, there is a dependency between values for each individual. With a LMM approach, it is possible to include each individual with his or her own "path" while still having a linear regression model. The model has been set up with HOOS₄ as the dependent variable, while timepoints, sex, age, Physical Component Summary, Mental Component Summary and BMI (the last four measured at baseline) as independent variables. For details about the covariance structure and other technical details, see the SPSS syntax (Additional File 4).

For the HOOS score, 1.54% of individual items were missing at Assessment A, 2.94% at B, 1.93% at C and 2.34% at D. A total score could be calculated for all subscales for 99% of the patients. Missing data were subsequently handled according to the HOOS scoring instructions [27] with the participate mean substitution method. Missing data were subsequently imputed with the mean of the other values within the same subscale. In order to quantify the difference in outcome change in respective HOOS subscales, a calculation of effect size was carried out. Effect size was computed as a mean score change divided by baseline SD (Kazis' effect size) [34].

For the comparison of the HOOS₄ scores at the four timepoints, marginal means have been computed. More specifically, this has been done with the implementation of such means in SPSS called “estimated marginal means”.

In order to determine whether the preoperative (baseline) evaluations were predictive of achieving either Functional Recovery or Treatment Failure, logistic regression models were created.

A number of independent continuous and categorical variables were examined. Patient-reported pain (from a visual analogue scale, VAS) was a continuous predictor. Subjects were divided according to sex and body mass index (BMI) and categorized as normal (<25), overweight (25–29.9) and obese (≥30). We also categorized individuals into tertiles based on their age (T1: <62.2 years, T2: 62.2–71.6 years, and T3: >71.6 years), scores in the Physical Component Summary (T1: <25.2, T2: 25.2–30.3 and T3: >30.3), and in the Mental Component Summary (T1: <35.9, T2: 35.9–44.3 and T3: >44.3) of the SF-36.

Missing data in all variables in regression analyses were imputed using the predicted value from an imputation model with the full conditional specification regression method for any missing pattern. Six imputed datasets were created, then the results were pooled in order to obtain a set of final parameter estimates that consequently were used in regression analyses [35].

Each independent variable was analyzed in an unadjusted and an adjusted model. The adjustment was made for sex, preoperative BMI, preoperative pain, Physical Component Summary and Mental Component Summary. The odds ratio (OR) estimates with 95% confidence intervals (95% CIs), and results from the likelihood ratio test, expressed as p values, were based on the models. Model fit was examined using the Hosmer-Lemeshow test [36].

All analyses were performed with the use of IBM SPSS Statistics for Windows v. 25.0.0 (IBM Corp. Armonk, New York, USA). We considered a two-tailed p value less than 0.05 to be significant. Sample size estimation was made with G*Power software v. 3.1.9.7 (Universität Düsseldorf, Germany) [37] (Additional File 1).

3. Results

Out of 195 subjects who underwent THR, 179 (92%) met the criteria and agreed to participate. At consecutive assessments, there were 153 (assessment B, 79%), 148 (C, 76%) and 140 patients (D, 72%) (Fig. 1). The study group consisted of 98 women (55%) and 81 men (mean age 67 (31–90), median 66 years). Women were significantly older than men (mean [SD] 70 [9] years, vs. 64 [11] years, p < 0.0001). One hundred and fifty-three patients (86%) were overweight or obese.

The patient characteristics are summarized in Table 1.

3.1. Patients lost to follow-up

Out of 179 patients who attended assessment A, 39 (22%) were lost to the final follow-up (assessment D). To evaluate a possible inclusion bias, the subjects who participated at assessment D and those who did not, were analyzed with regard to age, sex, BMI (Table 1), occupation activity

![Flowchart presenting the study group formation.](image-url)

Fig. 1. Flowchart presenting the study group formation.
and educational level (Additional File 3). We found no significant differences between these characteristics (data not shown).

3.2. Primary outcome

3.2.1. HOOS4

The mean HOOS4 scores improved substantially within first six weeks after THR (at assessment B) and continued to improve in consecutive assessments (C and D) \((p < 0.0001\) between following assessments). The score change between assessment A and D was 49.4 points producing a very large effect size of 2.47 (Fig. 2, Table 2).

3.3. Secondary outcomes

3.3.1. Change in the HOOS subscales

The differences in mean scores of all the HOOS subscales between consecutive assessments were statistically significant except for the difference in the HOOS subscale Sports and Recreation Function between assessments C and D. Score changes between assessment A and D exceeded 45 points for all subscales \((p < 0.0001, \text{effect sizes between } 2.93 \text{ for HOOS Symptoms and } 3.19 \text{ for the HOOS Pain})\) (Table 2).

![Fig. 2. Estimated marginal means of the HOOS4 scores at the four time points (A to D). Circle midpoints represent mean HOOS4 scores. Assessments were made at entry (preoperatively, assessment A), then 6–8 weeks after THR (B), 6 months after THR (C) and 12 months after THR (D). Vertical bars represent 95% confidence interval (95%CI) values. Asterisks (*) show } p \text{ values } < 0.001.\]

Table 1

| Characteristics                        | All subjects | Men  | Women |
|----------------------------------------|--------------|------|-------|
|                                        | n = 179      | n = 81 | n = 98 |
| Age, mean (SD), years                  | 66.9 (10.6)  | 63.7 (11.3) | 69.6 (9.3) |
| BMI, mean (SD)                         | 29.5 (4.4)   | 30.0 (4.3)   | 29.1 (4.4)   |
| BMI, n (%)                             | Normal 26 (15) | 8 (10)   | 18 (18)   |
|                                       | Overweight 81 (45) | 36 (44)   | 45 (46)   |
|                                       | Obese 72 (40) | 37 (46)   | 35 (36)   |
| SF-36, mean (SD)                       | Physical Component Summary 28.1 (6.4) | 29.2 (5.9)   | 27.3 (6.7)   |
|                                       | Mental Component Summary 39.5 (10.2) | 40.7 (10.0)   | 38.5 (10.4)   |

Abbreviations: BMI – body mass index; SD – standard deviation. SF-36 – 36-item Short Form Health Survey.

BMI was categorized as normal \((<25)\), overweight \((25–29.9)\) and obese \((\geq30)\). Percentage values were rounded to integers (whole numbers).

3.4. Pain

The mean preoperative VAS score (assessment A) was 6.83 (range 1–10, median 7). The pain score was higher in women than in men (mean [SD] 7.15 [1.61] vs. 6.44 [1.61], \(p = 0.002\)). Postoperatively, the VAS score was not remarkably different between assessments B and C but lower at D (mean 0.41 [1.07]). Postoperative differences between men and women were not significant \((p = 0.783 \text{ at assessment B}, 0.838 \text{ at C, and } 0.721 \text{ at D})\).

3.5. Functional Recovery

Functional Recovery was achieved in 41 out of 127 subjects (32%) who completed the HOOS questionnaire at assessment D. There was no significant difference in proportion of patients with Functional Recovery in men (14/57, 25%) and women (27/70, 39%) \((p = 0.127)\). All subjects with Functional Recovery improved between assessment A and D (data not shown).

Patients who had higher baseline scores \((>30.3, \text{tertile } 3)\) in the Physical Component Summary were more likely to achieve Functional Recovery than those who scored \(<25.2 \text{ (tertile } 1)\) in both models: unadjusted \((\text{OR} = 2.9; 95\% \text{CI } 1.1–7.5)\) and adjusted model \((\text{OR} = 3.2; 95\% \text{CI } 1.0–9.9)\). No association between age and Functional Recovery was observed (Table 3).

3.6. Treatment Failure

Treatment Failure was observed in 20 out of 127 (16%) subjects at assessment D. No difference in the proportion of men (10/57, 18%) and women (10/70, 14%) was observed \((p = 0.81)\). All patients with Treatment Failure improved in HOOS4 between assessment A and D.

Patients with preoperative Physical Component Summary values \(>30.3 \text{ (tertile } 3)\) were estimated to be less likely to have a Treatment Failure than those belonging to tertile 1 in both unadjusted \((\text{OR} = 1/0.067 = 14.9)\) and adjusted model \((\text{OR} = 1/0.097 = 10.3)\) (Table 4). Assuming that preoperative score \(>30.3\) is the reference level category, patients with values \(<25.2 \text{ (tertile } 1)\) were more likely to have Treatment Failure than those belonging to tertile 3 \((\text{OR} = 14.9, 95\% \text{CI } 1.8–121.6 \text{ in the unadjusted model and } \text{OR} = 10.3, 95\% \text{CI } 1.0–100.5 \text{ in the adjusted model})\). In addition, having preoperative Physical Component Summary score within tertile 2 \((25.2 < \text{PCS} < 30.3)\) was a factor associated with Treatment Failure in the unadjusted model \((\text{OR} = 9.9; 95\% \text{CI } 1.2–82.9)\).

Patients with preoperative Mental Component Summary values \(>44.3 \text{ (tertile } 3)\) were less likely to have a Treatment Failure than those with values \(<35.9 \text{ (OR} = 1/0.256 = 3.9)\) in an unadjusted model. Patients from tertile 2 (preoperative values between 35.9 and 44.3) were less likely to have a Treatment Failure than those belonging to tertile 1 in both unadjusted \((\text{OR} = 1/0.263 = 3.8)\) and adjusted model \((\text{OR} = 1/0.214 = 4.7)\) (Table 4). The other way around, preoperative Mental Component Summary \(<35.9\) was a factor significantly associated with confirming Treatment Failure in unadjusted \((\text{OR} = 3.9; 95\% \text{CI } 1.1–13.3)\) analysis.

4. Discussion

Our study aimed to prospectively document the improvement over 12 months in patients who had undergone THR as well as to assess Functional Recovery and Treatment Failure in these patients.

We found that patients after THR improved in both HOOS4 and in separate HOOS subscales. The mean difference between the baseline results (at assessment A) and the HOOS scores one year after THR (at assessment D) is nearly 50 points in the HOOS4, and even more than 50 points in two HOOS subscales: Symptoms and ADL. This remarkable improvement was higher than reported in other studies [38,39]. However, the preoperative scores in our study were between 14 points in
the HOOS subscale Sports and Recreation Function and 31 points in the subscale Pain, which was substantially lower than demonstrated in other studies.

Baseline HOOS results over 30 points in patients at similar age were reported by Nilsdotter et al. [33] and Arbab et al. [40]. In another study performed in slightly younger subjects who underwent THR with both short- and conventional stem implants, the baseline results of mean total HOOS score were 36 and 37 points, respectively, and over 25 points in the lowest-scored subscale Sport and Recreation Function [39].

The lower preoperative scores observed in our study could be explained by possible deterioration in physical function caused by longer waiting time for THR [41]. The time from referral to THR was 3 years in our study compared to less than 6 months in Denmark and 3 months in the Netherlands [42]. In addition, the mean BMI of our patients was higher than that reported by others [39,40].

As we observed, most of the improvement occurred in the first 6 months, with the steepest recovery trajectory within 6–8 weeks after THR. This observation corresponds to earlier findings, i.e. that most of the improvement occurs within the first 3 months [43,44]. According to the results of other studies, no [44] or only small clinical improvements [45,46] can be expected beyond the first 3-month period.

To the best of our knowledge, there have been no studies to compare outcome in patients undergoing THR with hip-specific population-based reference data. We have based the two concepts of Functional Recovery and Treatment Failure on the Swedish population study of Sundén et al. [21]. The Functional Recovery level, defined as the lower threshold for the 95% CI of sex- and age-matched subjects from population-based reference data, was described earlier for patients undergoing anterior cruciate ligament reconstruction and has been assessed with the Knee injury and Osteoarthritis Outcome Score (KOOS) [47].

Our study showed that despite an incremental score gain in the HOOS4 and all HOOS subscales, only 32% of patients who attended the final follow-up at assessment D achieved a Functional Recovery status, consequently attained the hip condition equal to or even better than average in age- and sex-matching population.

Outcomes such as pain and function are typically strongly associated with the level of patient satisfaction. Since satisfaction reflects also factors not directly related to the health status [48], it must be regarded as only an indirect or a proxy indicator of the intervention quality [49]. Since patient satisfaction has been reported to be highly dependent on preoperative patient expectations [50] and baseline scores [51], it is likely that a low preoperative score is associated with greater room for improvement and thus greater patient satisfaction [51]. Consequently, it is not surprising that patients from our cohort, who generally scored relatively low before THR, reported very high levels of satisfaction [28].

In the present study we use a definition of Treatment Failure based on the one-year follow-up score in the HOOS subscale QOL. Traditionally, failure of THR is defined as an implant-related problem demanding revision and requires a long-term observation. The modes of failure include implant wear, aseptic loosening, dislocation, infection, and periprosthetic fracture to name the most frequent ones [52]. Using our approach, we observed Treatment Failure in 16% of participants. This relatively high rate of unsatisfactory results can be associated with a potential mode of implant failure that appears later on, beyond one-year follow-up.

We did not observe any association of age and either Functional

---

**Table 2**

Scores in HOOS subscales and in VAS Pain in consecutive assessments: A (before THR), B (6–8 weeks after THR), C (6 months after THR) and D (12 months after THR).  

| HOOS Subscales | Assessment | A | B | C | D | p-values | ES (A vs. D) |
|---------------|------------|---|---|---|---|---------|------------|
| Pain          | Mean       | 31.2 | 70.7 | 77.2 | 80.3 | <0.001 | 0.001 | 0.046 | <0.001 | 3.19 |
|               | Median     | 30  | 70  | 82.5 | 85  |         | 0.001 | 0.001 | 0.177 |         |
|               | Range      | 0–82.5 | 22.5–100 | 27.5–100 | 22.5–100 | <0.001 | 0.001 | 0.003 | <0.001 | 2.93 |
| Symptom       | Mean       | 28.7 | 69  | 73.9 | 79.8 | <0.001 | 0.012 | 0.003 | <0.001 | 3.10 |
|               | Median     | 25  | 75  | 80  | 85  |         |         |         |         |         |
|               | SD         | 17.9 | 21.1 | 21.9 | 18.2 | <0.001 | 0.001 | 0.003 | <0.001 | 0.046 |
|               | Range      | 0–75 | 10–100 | 5–100 | 20–100 |         |         |         |         |         |
| ADL           | Mean       | 26.5 | 63  | 73.3 | 76.8 | <0.001 | <0.001 | <0.001 | <0.001 | 0.046 |
|               | Median     | 25  | 63  | 79  | 81  |         |         |         |         |         |
|               | SD         | 16.2 | 20.6 | 21.4 | 19.5 | <0.001 | 0.001 | 0.003 |         |         |
|               | Range      | 0–68 | 15–100 | 18–100 | 19–100 | <0.001 | 0.001 | 0.003 |         |         |
| Sport/Recreation | Mean       | 14.2 | 42.3 | 58  | 59.8 | <0.001 | 0.001 | 0.105 | <0.001 | 3.06 |
|               | Median     | 12.5 | 37.5 | 62.5 | 60.4 |         |         |         |         |         |
|               | SD         | 14  | 27  | 30  | 28.3 |         |         |         |         |         |
|               | Range      | 0–62.5 | 0–100 | 0–100 | 6–100 | <0.001 | <0.001 | 0.028 | <0.001 | 2.96 |
| QOL           | Mean       | 18   | 50.6 | 61.3 | 64.7 | <0.001 | <0.001 | 0.028 | <0.001 | 0.046 |
|               | Median     | 19   | 50   | 63  | 67  |         |         |         |         |         |
|               | SD         | 15.5 | 21.2 | 23.3 | 22.3 | <0.001 |         |         |         |         |
|               | Range      | 0–63 | 6–94  | 0–100 | 6–100 | <0.001 | <0.001 | <0.001 | <0.001 | 2.47 |
| HOOS4         | Mean       | 26   | 63.2 | 71.5 | 75.4 | <0.001 | <0.001 | <0.001 | <0.001 | 0.046 |
|               | Median     | 26.2 | 63.2 | 76.1 | 79.6 |         |         |         |         |         |
|               | SD         | 14.4 | 18.5 | 20.2 | 18  | <0.001 |         |         |         |         |
|               | Range      | 0–67 | 18–97.5 | 18–100 | 19–100 | <0.001 | <0.001 | <0.001 | <0.001 | 2.47 |
| VAS Pain      | Mean       | 6.8  | 0.9  | 0.8  | 0.4  | <0.001 | 0.197 | 0.015 | <0.001 | 3.25 |
|               | Median     | 7    | 0    | 0    | 0    |         |         |         |         |         |
|               | SD         | 1.6  | 1.5  | 1.5  | 1.1  |         |         |         |         |         |
|               | Range      | 1–10 | 0–7  | 0–7  | 0–5  | <0.001 |         |         |         |         |

Abbreviations: HOOS – Hip disability and Osteoarthritis Outcome Score; THR – total hip replacement; ADL – activities of daily living; QOL – quality of life; ES – effect size.

P-values for the comparison between the results in consecutive assessments: A and B, B and C, C and D as well as A and D. Kazi’s effect size (ES) was used for the comparison between results in assessment A and D.
Odds ratios for functional recovery after THR (at assessment D) by patient characteristics and selected predictors at assessment A.

| Factor | Number of cases | Unadjusted model | Adjusted model |
|--------|-----------------|------------------|---------------|
|        | n   | %  | OR  | 95% CI | OR  | 95% CI |
| Sex    |     |    |     |        |     |        |
| Man    | 57  | 45 | 1.9 | 0.9-4.2 | 2.4 | 1.0-6.0 |
| Woman  | 70  | 55 | 1   | 1      | 1   | 1      |
| Age    |     |    |     |        |     |        |
| T1a    | 42  | 37 | 2.2 | 0.9-5.6 | 2.2 | 0.8-6.6 |
| T2     | 43  | 31 | 1.5 | 0.6-4.0 | 1.6 | 0.5-4.6 |
| T3     | 42  | 32 | 2.2 | 0.9-5.6 | 2.2 | 0.8-6.6 |
| BMI    |     |    |     |        |     |        |
| Normal | 21  | 17 | 1.2 | 0.4-3.6 | 0.6 | 0.2-2.2 |
| Overweight | 55  | 43 | 1.3 | 0.6-2.9 | 1.0 | 0.4-2.5 |
| Obese  | 51  | 40 |    |        |     |        |
| PCS    |     |    |     |        |     |        |
| T1a    | 42  | 33 | 1.6 | 0.6-4.4 | 1.7 | 0.6-5.0 |
| T2     | 43  | 34 | 2.9 | 1.1-7.5 | 3.2 | 1.0-9.9 |
| T3     | 42  | 33 | 1.0 | 0.4-2.6 | 0.9 | 0.3-2.7 |
| VAS    |     |    |     |        |     |        |
| 125 100 | 0.9 | 0.7-1.1 | 1.0 | 0.7-1.2 |

**Abbreviations:** THR – total hip replacement; BMI – body mass index; PCS – physical component summary; MCS – mental component summary; OR – odds ratio; 95% CI – 95% confidence interval; T – tertile.

Subjects were categorized into tertiles according to the distribution of age (T1: <62.2 years, T2: 62.2–71.6 years, and T3: >71.6 years) to the baseline PCS scores (T1: <25.2, T2: 25.2–30.3 and T3: >30.3) and MCS scores (T1: <35.9, T2: 35.9–44.3 and T3: >44.3). OR and 95% CI values were rounded to decimals.

Odds ratios for Treatment Failure after THR (at assessment D) by patient characteristics and selected predictors at assessment A.

| Factor | Number of cases | Unadjusted model | Adjusted model |
|--------|-----------------|------------------|---------------|
|        | n   | %  | OR  | 95% CI | OR  | 95% CI |
| Sex    |     |    |     |        |     |        |
| Man    | 57  | 45 | 0.8 | 0.3-2.0 | 0.3 | 0.1-1.0 |
| Woman  | 70  | 55 | 1   | 1      | 1   | 1      |
| Age    |     |    |     |        |     |        |
| T1a    | 42  | 37 | 0.3 | 0.1-1.3 | 0.3 | 0.1-1.5 |
| T2     | 43  | 31 | 1.2 | 0.4-3.4 | 1.4 | 0.4-5.2 |
| T3     | 42  | 32 | 1.7 | 0.5-5.9 | 3.0 | 0.6-14.5 |
| BMI    |     |    |     |        |     |        |
| Normal | 21  | 17 | 0.8 | 0.3-2.3 | 0.8 | 0.2-2.7 |
| Overweight | 55  | 43 | 0.3 | 0.1-0.9 | 0.2 | 0.1-0.9 |
| Obese  | 51  | 40 |    |        |     |        |
| PCS    |     |    |     |        |     |        |
| T1a    | 42  | 33 | 0.7 | 0.2-1.9 | 0.7 | 0.2-2.2 |
| T2     | 42  | 33 | 0.1 | 0-0.6  | 0.1 | 0-1.0  |
| T3     | 43  | 34 | 0.3 | 0.1-0.9 | 0.3 | 0.1-1.2 |
| VAS    |     |    |     |        |     |        |
| 125 100 | 1.2 | 0.9-1.2 | 1.1 | 0.8-1.7 |

**Abbreviations:** THR – total hip replacement; BMI – body mass index; PCS – physical component summary; MCS – mental component summary; OR – odds ratio; 95% CI – 95% confidence interval; T – tertile.

BMI was categorized as normal (<25), overweight (25–29.9) and obese (≥30). Subjects were categorized into tertiles according to the distribution of age (T1: <62.2 years, T2: 62.2–71.6 years, and T3: >71.6 years) to the baseline PCS scores (T1: <25.2, T2: 25.2–30.3 and T3: >30.3) and MCS scores (T1: <35.9, T2: 35.9–44.3 and T3: >44.3). OR and 95% CI values were rounded to decimals.

A slight female predominance with the female/male ratio of 1.2 and the higher average age of women undergoing THR reflects a well-known observation that women with OA undergo surgery at a later stage of the disease [8]. Despite discrete ethnic and cultural differences, we believe that the reference data from the Swedish population can even be applied in Poland. Since the sample size was relatively small, missing data, even if there were only few, reduced the power of our analyses. Missing values at the item level have been imputed successfully by mean substitution, a method which has its drawbacks, mainly that correlations between imputed variables are attenuated. However, this method was chosen because it was regarded as important to use the same protocol for the calculation of the HOOS scales as suggested by the original authors. As we assumed that the incomplete data at the variable level (for those covariates that were included in regression analyses) were missing at random, we have replaced them by multiple imputation using the full conditional specification method. This strategy allowed us to correct bias and substantially improve precision of calculations.

Based on data concerning objective postoperative evaluations and patient improvement observed in PROMs, clinicians can have an over-rated picture of patients’ health after THR. Since the results of THR are associated with fulfillment of patient preoperative expectations (irrespective of whether they are high or low), setting the reasonable goals of treatment seems to be essential to achieve optimal postoperative results.
[59]. Knowing that only one third of subjects undergoing THR will reach population-based Functional Recovery, clinicians will have a better perspective to discuss rationality of expectations and, consequently, prevent patient dissatisfaction.

Authors’ contributions

PTP conceived and designed the study. MKG was responsible for acquisition of data and creation of datasets. PTP and MKG drafted the manuscript. RL takes responsibility for the integrity of the data and the accuracy of the data analysis. The article was revised critically for important intellectual content by PTP. All authors participated in the analysis and interpretation of the data, and approved the final version for submission.

Role of the funding source

This work was supported by the Faculty of Health Sciences, Ludwik Rydygier Collegium Medicum of the Nicolaus Copernicus University, Toruń, Poland (grant WN 758), and by the County Council of Norrbotten, Sweden.

Ethics approval and consent to participate

The study was conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The patients were informed in writing and orally by the study personnel, and a written informed consent was obtained from all subjects. Participation was voluntary, and withdrawal was possible at any time. All patients signed and personally dated the informed consent forms at admission to hospital, before participating in the study. The study was approved by the Medical Ethics Committee of the Local Chamber of Physicians and Dentists of the Warmia and Mazury Province in Olsztyn (Approval no. 140/2013).

Declaration of competing interests

The authors declare no conflict of interests.

Acknowledgements

The authors would like to personally acknowledge Professor Martin Englund for his help and consistent support and Robert Foltyn for the assistance in preparing the manuscript.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2021.100192.

APPENDIX A. ADDITIONAL FILES

The following are the supplementary data to this article:

1. Sample size estimation and power analysis
2. Surgical procedure and postoperative management
3. Occupation activity and educational level
4. SPSS syntax for the HOOS modelling

References

[1] I.D. Learmonth, C. Young, C. Coraback, The operation of the century: total hip replacement, Lancet 370 (2007) 1508–1519, https://doi.org/10.1016/S0140-6736(07)60457-7.
[2] R. Pivec, A.J. Johnson, S.C. Mears, M.A. Mont, Hip arthroplasty, Lancet 380 (2012) 1768–1777, https://doi.org/10.1016/S0140-6736(12)60607-2.
[3] S.J. Birtwistle, K. Wilson, M.I. Porter, Long-term results of total hip replacement, Ann. R. Coll. Surg. Engl. 78 (1996) 180–183, https://www.ncbi.nlm.nih.gov/pubmed/8779498.
[4] C. Varum, Outcomes of different bearings in total hip arthroplasty - implant survival, revision causes, and patient-reported outcome, Dan. Med. J. (2017) 64. https://www.ncbi.nlm.nih.gov/pubmed/28260601.
[5] J.T. Evans, J.P. Evans, R.W. Walker, A.W. Blom, M.R. Whitehouse, A. Sayers, How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up, Lancet 393 (2019) 647–654, https://doi.org/10.1016/S0140-6736(18)31665-9.
[6] L. Nikolasjen, B. Brandsborg, U. Lucht, T.S. Jensen, H. Kehlet, Chronic pain following total hip arthroplasty: a nationwide questionnaire study, Acta Anaesthesiol. Scand. 50 (2006) 495–500, https://doi.org/10.1111/j.1399-6576.2006.00976.x.
[7] R.E. Anakwe, P.J. Jenkins, M. Moran, Predicting dissatisfaction after total hip arthroplasty: a study of 850 patients, J. Arthroplasty 26 (2011) 209–213, https://doi.org/10.1016/j.arthro.2010.03.013.
[8] Swedish hip arthroplasty register. Annual Report 2015, Svenska Hjärtprostregistreret, Gothenburg, Sweden, 1979–2018, https://registercentrum.blob.core.windows.net/ocarto/2015-280116D7C2013Pc.pdf.
[9] P.C. Noble, S. Fuller-Lafreniere, M. Mefait, M.K. Dwyer, Challenges in outcome measurement: discrepancies between patient and provider definitions of success, Clin. Orthop. Relat. Res. 471 (2013) 3437–3445, https://doi.org/10.1199/10.1199/3199-01-3196-2.
[10] M.J. Halawi, Outcome measures in total joint arthroplasty: current status, challenges, and future directions, Orthopedics 38 (2015) e685–689, https://doi.org/10.3928/01477447-20151004-55.
[11] J. Wilson, E. Bohm, A. Lubbeke, S. Lyman, S. Overgaard, O. Rolfson, et al., Orthopaedic registries with patient-reported outcome measures, EFORT Open Rev. 4 (2019) 357–367, https://doi.org/10.1016/j.efortoprev.2019.08.002.
[12] S.N. Hofstede, M.G. Gademan, T.P. Vliet Vlieland, R.G. Nelissen, P.J. Marang-van de Merode, Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review, BMJ Musculos. Disord. 17 (2016) 212, https://doi.org/10.1186/s12911-016-1070-3.
[13] Z.H. Dailiana, I. Papakostidou, S. Varitimidis, L. Liaropoulos, E. Zintzaras, T. Karachalios, et al., Patient-reported quality of life after primary major joint arthroplasty: a prospective comparison of hip and knee arthroplasty, BMJ Musculos. Disord. 16 (2015) 366, https://doi.org/10.1186/s12911-015-0814-9.
[14] A.K. Nilsson, F. Isaksson, Patient relevant outcome 7 years after total hip replacement for OA - a prospective study, BMJ Musculos. Disord. 11 (2010) 47, https://doi.org/10.1186/1471-2474-11-47.
[15] P.J. van der Wees, J.J. Wamms, R.P. Akkermans, J. Koetsenruijter, G.P. Westert, A. van Kampen, et al., Patient-reported health outcomes after total hip and knee surgery in a Dutch University Hospital Setting: results of twenty years clinical registry, BJM Musculos. Disord. 18 (2017) 97, https://doi.org/10.1186/s12911-017-1455-y.
[16] M.M. Vissers, J.B. Bussmann, I.B. de Groot, J.A. Verhaar, M. Beijman, Physical functioning four years after total hip and knee arthroplasty, Gait Posture 38 (2013) 310–315, https://doi.org/10.1016/j.gaitpost.2012.12.007.
[17] N.K. Arden, A. Kiran, A. Judge, L.C. Biant, M.K. Javaid, D.W. Murray, et al., What is a good patient reported outcome after total hip replacement? Osteoarthritis Cartilage 19 (2011) 155–162, https://doi.org/10.1016/j.joca.2010.00.004.
[18] R.M. Peters, L. van Beers, L.N. van Steenbergen, J. Wolkentfeld, H.B. Ettema, B. Ten Have, et al., Similar superior patient-reported outcome measures for anterior and posterolateral approaches after total hip arthroplasty: postoperative patient-reported outcome measure improvement after 3 months in 12,774 primary total hip arthroplasties using the anterior, anterolateral, straight lateral, or posterolateral approach, J. Arthroplasty 33 (2018) 1786–1793, https://doi.org/10.1016/j.arth.2018.01.055.
[19] J.M. McLean, J. Cappelletto, J. Clamette, C.L. Hill, T. Gill, D. Mandziuk, et al., Normal population reference values for the Oxford and Harris Hip Scores - electronic data collection and its implications for clinical practice, Hip Int. 27 (2017) 389–396, https://doi.org/10.1186/s11999-017-0220-9.
[20] A. Sundén, K. Lidgren, E.M. Ross, L.S. Lohnander, E. Ekvall Hansson, Hip complaints differ across age and sex: a population-based reference data for the hip disability and Osteoarthritis Outcome Score (HOOS), Health Qual. Life Outcome 16 (2018) 200, https://doi.org/10.1186/s12955-018-1022-8.
[21] J. C Zhuang, D. Goggan, I. Reading, P. Grob, P. Byng, K. Cox, et al., Long-term outcome following total hip arthroplasty: a controlled longitudinal study, Arthritis Rheum. 57 (2007) 1375–1380, https://doi.org/10.1002/art.3101.
[22] A. Judge, C. Cooper, N.K. Arden, S. Williams, N. Holbo, D. Dixon, et al., Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres, Osteoarthritis Cartilage 19 (2011) 659–667, https://doi.org/10.1016/j.joca.2011.03.009.
[23] A. Judge, M.K. Javaid, N.K. Arden, J. Cusnaghan, I. Reading, P. Croft, et al., Clinical tool to identify patients who are most likely to achieve long-term improvement in physical function after total hip arthroplasty, Arthritis Care Res. 64 (2012) 881–889, https://doi.org/10.1002/acr.21594.
[24] P.R. Fortin, J.R. Penrod, A.E. Clarke, Y. St-Pierre, L. Joseph, P. Belisle, et al., Timing of total joint replacement affects clinical outcomes among patients with
ostearthritis of the hip or knee, Arthritis Rheum. 46 (2002) 3327–3330, https://doi.org/10.1002/art.10631.

[26] R. Alman, G. Alarcon, D. Appelrouth, D. Bloch, D. Borenstein, K. Brandt, et al., The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip, Arthritis Rheum. 34 (1991) 505–514, https://doi.org/10.1002/art.1780340502.

[27] Hip Disability and Osteoarthritis Outcome Score, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, 2012-2016. http://www.koos.nu/koospolish.pdf.

[28] M.K. Gojo, P.T. Paradossi, Polish adaptation and validation of the hip disability and osteoarthritis outcome score (HOOS) in osteoarthritis patients undergoing total hip replacement, Health Qual. Life Outcome 18 (2020) 135, https://doi.org/10.1186/s12955-020-01940-3.

[29] D. Žolnierzczyk-Zreda, The Polish version of the SF-36v2 questionnaire for the quality of life assessment, Przegl. Lek. 67 (2013) 1302–1307.

[30] N.C. Laucis, R.D. Hays, T. Bhattacharyya, Scoring the SF-36 in orthopedics: a brief guide, J. Bone Joint. Surg. Am. 97 (2015) 1628–1634, https://doi.org/10.2106/JBOS.00030.

[31] S.S. Fafariv, W.L. Cunningham, R.D. Hays, Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I, Health Qual. Life Outcome 5 (2007) 54, https://doi.org/10.1186/1477-7525-5-54.

[32] J. Scott, E.C. Huskisson, Graphic representation of pain, Pain 2 (1976) 175–184, https://www.ncbi.nlm.nih.gov/published/10269600.

[33] A.K. Nilsdotter, L.S. Lohmander, M. Klassbo, E.M. Roos, Hip disability and function over 5 years following primary hip-replacement surgery: a prospective blinded randomised multicentre trial comparing the Collum Femoris Preserving stem with a Zweymuller straight cementless stem total hip replacement, BMC Musculoskelet. Disord. 15 (2014) 53, https://doi.org/10.1186/1471-2474-15-53.

[34] L.E. Kazis, J.J. Anderson, R.F. Meenan, Effect sizes for interpreting changes in health status, Med. Care 27 (1989) S178–S189, https://doi.org/10.1097/00005482-19890300-00015.

[35] S.R. Seaman, R.H. Keogh, Handling missing data in matched case-control studies using multiple imputation, Biometrics 71 (2015) 3–15, https://doi.org/10.1111/biom.12358.

[36] D.W. Hosmer, T. Hosmer, S. Le Cessie, S. Lemeshow, A comparison of goodness-of-fit tests for the logistic regression model, Stat. Med. 16 (1997) 965–980, https://doi.org/10.1002/(sici)1097-0258(19970515)16:9<965::aid-sim509>3.0.co;2-e.

[37] F. Paul, E. Erdtelder, A. Buchner, A.G. Lang, Statistical power analyses using G’Power 3.1: tests for correlation and regression analyses, Behav. Res. Methods 41 (2009) 1149–1160, https://doi.org/10.3758/BRM.41.4.1149.

[38] S. Rosendlund, L. Broeng, A. Holgaard-Larsen, C. Jensen, S. Overgaard, Patient-reported outcome after total hip arthroplasty: comparison between lateral and posterior approach, Acta Orthop. 88 (2017) 239–247, https://doi.org/10.1080/19477058.2016.1271910.

[39] J. van Oldenrijk, V.A.B. Scholtes, L. van Beers, C.H. Geerdink, B. Niers, W. Runne, et al., Early clinically relevant improvement in quality of life and clinical outcomes 1 Year postsurgery in patients with knee and hip joint arthroplasties, Cartilage 9 (2018) 127–139, https://doi.org/10.1177/1947605177439000.

[40] B. Barenius, M. Forssblad, B. Engstrom, K. Eriksson, Functional recovery after anterior cruciate ligament reconstruction, a study of health-related quality of life based on the Swedish National Knee Ligament Register, Knee Surg. Sports Traumatol. Arthrosc. 21 (2013) 914–927, https://doi.org/10.1007/s00167-012-2162-8.

[41] A.R. Tyser, C.J. Gaffney, C. Zhang, A.P. Presson, The association of patient satisfaction with pain, anxiety, and self-reported physical function, J. Bone Joint. Surg. Am. 100 (2018) 1811–1818, https://doi.org/10.2106/JBJS.17.00372.

[42] B. Prakash, Patient satisfaction, J. Cutan. Aesthetic Surg. 3 (2010) 151–155, https://doi.org/10.4103/0974-2077.74491.

[43] C. Palazzo, C. Jourdan, S. Descamps, R. Nizard, M. Hamadouche, P. Anract, et al., Determinants of satisfaction 1 year after total hip arthroplasty: the role of expectations fulfillment, BMC Musculoskelet. Disord. 15 (2014) 53, https://doi.org/10.1186/1471-2474-15-53.

[44] A. Paulsen, E.M. Roos, A.B. Pedersen, S. Overgaard, Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively, Acta Orthop. 85 (2014) 39–48, https://doi.org/10.3109/17456367.2013.867782.

[45] T. Karachalios, G. Kommos, A. Koulatsis, Total hip arthroplasty: survival and modes of failure, EORTC Open Rev. 3 (2018) 232–239, https://doi.org/10.1302/2058-5241.3.170066.

[46] R. Gandhi, H. Dhotar, J.R. Davey, N.M. Mahomed, Predicting the longer-term outcomes of total hip replacement, J. Rheumatol. 37 (2010) 2573–2077, https://doi.org/10.2899/jrheum.100149.

[47] A. Judge, N.K. Arden, R.N. Batra, G. Thomas, D. Beard, M.K. Javad, et al., The association of patient characteristics and surgical variables on symptoms of pain and function over 5 years following primary hip-replacement surgery: a prospective cohort study, BMJ Open 3 (2013), https://doi.org/10.1136/bmjopen-2012-002453.

[48] M. Gordon, M. Greene, P. Fronmento, O. Rolfson, G. Garelick, A. Stark, Age and health-related quality of life after total hip replacement: decreasing gains in patients above 70 years of age, Acta Orthop. 85 (2014) 244–249, https://doi.org/10.3109/17456367.2014.916492.

[49] M. Weber, F. Zeman, B. Croaiovan, M. Thiemes, M. Kaiser, M. Woerner, et al., Predicting outcome after total hip arthroplasty: the role of preoperative patient-reported measures, BioMed Res. Int. (2019) 4909561, https://doi.org/10.1155/2019/4909561.

[50] J.L. Berliner, D.J. Brodkin, V. Chan, N.F. SooHoo, K.J. Bozic, John Charnley Award, Preoperative patient-reported outcome measures predict clinically meaningful improvement in function after THA, Clin. Orthop. Relat. Res. 474 (2016) 321–329, https://doi.org/10.1007/s11999-015-4590-6.

[51] T. Duivenvoorden, M.M. Vissers, J.A. Verhaar, J.J. Busschbach, T. Gosens, R.M. Bloem, et al., Anxiety and depressive symptoms before and after total hip and knee arthroplasty: a prospective multicentre study, Osteoarthritis Cartilage 21 (2013) 1834–1840, https://doi.org/10.1016/j.joca.2013.08.022.

[52] C.A. Mancuso, E.A. Salvati, N.A. Johanson, M.G. Peterson, M.E. Charlson, Patients’ expectations and satisfaction with total hip arthroplasty, J. Arthroplasty 12 (1997) 387–396, https://doi.org/10.1080/08835407909701947.