Can the Minimal Clinically Important Difference be determined in a French-speaking population with primary hip replacement using one PROM item and the Anchor strategy?

Sophie Putman, Julien Dartus, Henri Migaud, Gilles Pasquier, Julien Girard, Cristian Preda, Alain Duhamel

Original article

Can the Minimal Clinically Important Difference be determined in a French-speaking population with primary hip replacement using one PROM item and the Anchor strategy?

Sophie **Putman** [a,b,c,d*], Julien **Dartus** [a,b], Henri **Migaud** [a,b], Gilles **Pasquier** [a,b], Julien **Girard** [a,b], Cristian **Preda** [b,e,f], Alain **Duhamel** [b,c,d]

a CHU Lille, Service d'Orthopédie, Hôpital Salengro, Place de Verdun, F-59000 Lille, France
b Université Lille, 59000, France
c Univ. Lille, CHU Lille, ULR2694 –METRICS : Evaluation des Technologies de Santé et des Pratiques Médicales, F-59000 Lille, France
d CHU Lille, Department of Biostatistics, F-59000 Lille, France
e Laboratory of Mathematics Paul Painlevé, UMR CNRS 8524, University of Lille, France
f Lille Catholic Hospitals, Biostatistics Department, Delegation for Clinical Research and Innovation, Lille Catholic University, Lille, France

* **Corresponding author**: Sophie Putman, CHU Lille, Service d'Orthopédie, Hôpital Salengro, Place de Verdun, F-59000 Lille, France
E-mail: sophie.putman@wanadoo.fr

Abstract

Background

The impact of surgery on the patient is classically assessed on pre- and post-treatment scores. However, it is increasingly recommended to rank these results according to the Minimal Clinically Important Difference (MCID), using either the data distribution method or the anchor method, latter consisting in an extra question specifically targeting the patient's improvement. MCIDs vary between populations and to the best of our knowledge there have been no investigations in France regarding this in the context of total hip replacement (THR). Therefore, we conducted a prospective study in a population with THR to determine: 1) whether MCID scores in France were comparable to those reported in the data from th international literature; 2) whether a general item taken from a different score could serve as an

anchor; and 3) whether an item from the actual questionnaire itself could serve as an anchor.

Hypothesis

When pre- and post-treatment scores are available, an item from the questionnaire itself can serve as an anchor for MCID.

Material and methods

In a prospective observational study, 123 primary THR patients (69 male, 54 female), out of 150 initially included, completed the 5 domains of the HOOS hip disability and osteoarthritis outcome score and the Oxford-12 questionnaire, preoperatively and at 6-12 months. The MCID was calculated via the distribution-based and the anchor-based methods. Two Oxford items (questions 1 and 2) and 2 HOOS items (questions S1 and Q4) were used as anchors, as well as a supplementary question on improvement and the Forgotten Joint Score (FJS).

Results

At a mean 10.12±1.2 months' follow-up [range, 6.5-11.9 months], the Oxford-12 score increased from 19±8 [3-35] to 40±10 [8-48] ($p<0.001$), all HOOS components demonstrated improvement, and the FJS at the final follow-up was 71±29 [0-100]. The general items (Oxford question 1 and HOOS question Q4) were more discriminating than the joint-specific items (Oxford question 2 and HOOS question S1). Based on results from the 3 anchors (improvement rated 1 to 5, Oxford question 1 and HOOS question Q4), 3 to 5 patients showed deterioration, 5 to 6 were unchanged, 30 to 40 were slightly improved, and 73 to 80 were improved by THR. The mean MCID on both distribution and anchor methods was 9 [5.5-12] for 'Oxford-12, 20 [12-27] for HOOS Symptoms, 26 [10-36] for HOOS Pain, 22 [11.5-28] for HOOS Function, 26 [13-34] for HOOS Sport and 22 [14-28] for HOOS Quality of Life.

Discussion

The MCID for the Oxford-12 and HOOS scores in a French population was comparable to data from the past literature. Using a score item as an anchor to define improvement is possible, but only if a general item is used.

Level of evidence: IV; prospective study without control group

Keywords: PROM, Hip, HOOS, Total hip replacement

Clinical Trials registration n°: NCT04057651

1. Introduction

Patient-Reported Outcome Measures (PROM) are increasingly used in postoperative assessment, and they themselves need to be assessed [1]. Their utility is to provide scores for patient status to be evaluated. With their increasing use, Jaesche et al. [1] introduced the concept of the Minimal Clinically Important Difference (MCID), being the smallest difference patients consider important. Although the MCID differs between populations and according to the method by which it is determined, it is increasingly used in the context of questionnaire assessment [2,3] and is important for interpreting the results of surgery and differences between procedures. In trials using PROMs as the main endpoint, the MCID determines the expected difference in superiority studies or the non-inferiority threshold in non-inferiority studies.

There are two methods [4,5] most commonly used to calculate the MCID: one based on score distribution, and the other using an extra question known as an “anchor” to categorize results as “improvement/no change/worsening”. In a systematic review of the literature, Celik et al. [6] analyzed this concept in questionnaires dedicated to the hip, knee and ankle joints, and found a variety of methods for choosing an anchor. Some scores, such as the Hip Osteoarthritis Outcome Score (HOOS) [7,8], include a general item: “Finally, are you bothered by your hip?”, with 5 possible responses. In patient assessment, these scores are available pre- and post-treatment, including the answers to these general questions that help in interpreting the results.

The MCID varies between populations [9] and, to our knowledge, there have been no French investigations in populations with total hip replacement (THR). Therefore, we conducted a prospective study in a primary THR population, using the HOOS [7-8] and Oxford-12 [10-11] scores, to determine: 1) whether the MCIDs for these scores in France were comparable to the international literature data; 2) whether a general item taken from a different score could serve as an anchor; and 3) whether an item from the actual questionnaire itself could serve as an anchor. We hypothesized that, when pre- and post-treatment scores are available, an item from the questionnaire itself can serve as anchor for MCID.

2. Material and methods

2.1 Patients

A prospective study was conducted between 2017 and 2019 in the University Hospital Center of Lille, France (local registration n° 2017-A01911-52; Clinical Trials registration n° NCT04057651) with approval from the “Est IV” data protection

committee. All patients aged ≥18 years undergoing THR for osteoarthritis or osteonecrosis were included on signature of informed consent.
Exclusion criteria comprised: lack of French national health insurance cover, age <18 years, inability to provide consent, pregnancy, and body-mass index > 35 or <18. Thus, 150 patients were included: 90 male, 60 female; mean age 64 ± 13 years (range, 34-89 years).

2.2 Methods

Patients completed the hip-specific HOOS and Oxford-12 questionnaires [7, 8,10,11] on the eve of surgery, and then via mail at 6-12 months, when the FJS Forgotten Joint Score was also associated [12].
They also responded to a question rating their perceived level of improvement; "Compared to before surgery, how do you rate the present state of your hip? (1 = worse, 2 = slightly worse, 3 = the same, 4 = slightly better, 5 = better or much better)". 123 of the 150 patients (82%: 69 male, 54 female) responded at ≥6 months.

2.3 Assessment

Two methods were used to determine the MCID [4-6,13]:

1) Distribution-based

This is a statistical method based on the distribution of the variable over the population as a whole, determining the minimal change required for the response to be better than chance. In the overall population, we used the mean pre- to post-treatment change, Δ, with standard deviation (SD), such that SD_{Δ} *0.5 = MCID [14].
We also used the Minimal Detectable Change (MDC) at 95% confidence level: MDC95 = 1.96$\sqrt{2}$SEM (Standard Error of Measurement) [15].

2) Anchor-based

Several anchors were tested, with the 5-point improvement question as a reference.
For the HOOS, the anchor was question 4: "In general, how much difficulty do you have with your hip?" with 5 possible responses (Figure 1).
For Oxford-12, the anchor was question 1: "How would you describe the pain you usually have in your hip?", again with 5 responses (Figure 2)

To assess discrimination, we applied the same procedure on 2 items unrelated to clinical improvement:

- Oxford item 2: “Have you had any trouble with washing and drying yourself (all over) because of your hip?”;
- and HOOS item S1: “Do you feel grinding, hear clicking or any other kind of noise from your hip?”

For these anchors, results were considered unchanged if pre- and post-operative responses were the same (0), and improved for an improvement of ≥1 point (Figures 1 and 2). Improvement of 1 point was taken as equivalent to the response “slightly better” and improvement of 2 points to “better or much better”.

The ROC was used to assess the discriminatory power of pre- to post-operative difference in scores in 2 sub-populations: with and without improvement. The group without improvement comprised patients with worsened, unchanged and only slightly improved results; the group with improvement comprised patients responding “better or much better”.

Areas Under the ROC (receiver operating curve) Curve (AUC) were calculated. The score sensitivity corresponded to the percentage of patients showing improvement in agreement with the anchor, with the score differential exceeding the MCID. The specificity corresponded to the percentage without improvement in agreement with the anchor, with the score differential less than the MCID. Thus, the largest possible AUC was sought and values >0.9 were considered exceptional, 0.8-0.9 excellent, 0.7-0.8 acceptable and 0.5 non-discriminating [16]. The MCID was calculated for anchors that were more discriminating than the reference anchor (the 1-5 point improvement question).

Once populations had been determined for each anchor considered as discriminating, the Younden index was used to determine the MCID threshold: i.e., point on the ROC curve where [sensitivity + specificity -1] was optimal [17].

2.4 Statistics

Statistical analysis was conducted using SPSS software (IBM, Bois-Colombes, France). Quantitative variables were reported as the mean and standard deviation for normal distribution or else the median and interquartile range. Distribution normality was checked graphically and assessed on the Shapiro-Wilk test. Qualitative variables were reported as numbers and percentages, and these were compared between groups on Student test, after checking normal distribution. The significance threshold was set at p=0.05. The effect size was calculated using Cohen’s criteria: 0.2, weak;

0.5, moderate; and 0.8, strong [18].

3. Results

3.1 Score distributions

123 of the initial 150 patients (69 male, 54 female) responded, at a mean of 10.12±1.2 months (range, 6.5-11.9 months). Table 1 shows the preoperative and follow-up scores. There was significant systematic overall improvement ($p < 0.001$).

3.2 Choice of anchors

Table 2 shows the AUCs. They lay between 0.7 and 0.9 for the 3 “general” anchors: improvement question, Oxford question 1 and HOOS question Q4. The AUCs for the Oxford question 1 and HOOS question Q4 were slightly greater than for the usual 1-5 point improvement question, and could thus be considered discriminating and be used for calculating the MCID.

In contrast, the anchors with little relevance to improvement (Oxford question 2 and HOOS question S1) were less discriminating than the general items (Oxford question 1 and HOOS question Q4), with smaller AUCs, close to 0.5, indicating that they could not be used for calculating the MCID (Figure 3).

3.3 Score distributions and calculation of MCID according to anchor

Using the 3 anchors (improvement question, Oxford question 1 and HOOS question Q4), there were 3-5 patients with deterioration, 5-6 unchanged, 30-40 slightly improved, and 73-80 improved by surgery (Table 3). Thus, there was an overall improvement except in 9-10 patients, who were dissatisfied or unchanged, this number being comparable between anchors. The MCID was calculated according to the predefined distribution-based method and the method based on the most discriminating anchors (Table 4).

4. Discussion

4.1 Results for the series

This study showed that the MCID in a French-speaking population was similar to reports for other populations (Tables 5 and 6). These findings were borne out by an

FJS-12 of 71±29 points, close to the mean value of 70.9±33 reported by Giesinger et al. in a general US population [26].

Even so, although the study defined threshold values for a French population for the first time, the results varied according to method and are only indicative. Results vary between countries [23] and Lyman et al. [3] suggested that environmental factors could explain the difference found in Denmark [23], with better quality of life on international rankings. Results may also differ between populations according to the findings of Kuo et al. [24], suggesting specific results in a population of veterans. The MCID may also vary with pathology as values after hip arthroscopy varied slightly, notably for the HOOS (Table 6) [21].

Results vary further according to the method. Kuo et al. [25] showed that, like in the present study, the distribution-based method gave smaller MCIDs than the anchor-based method. For the distribution-based method, Copay et al. [5] recommend using the MDC95, to have a result independent of sample size. The ROC curve analysis is also debatable, as it depends on the choice of index with some groups using the Youden index and others using a specificity value of 0.8 [27,28].

4.2 Can an item from one score serve as anchor for another?

For an item from one score to serve as an anchor for another, it has to be general enough to assess progression. Less general items were less sensitive to improvement than items focusing on quality of life or pain. Question 1 of the Oxford-12 score, for example, was interesting, as Goodman et al. [24] had shown preoperative pain to be related to improvement. Danoff et al. [29] defined an MCID by pain on the visual analogue scale. Other studies also used items from different scores: Lyman et al. [3] used the Satisfaction Survey of the HSS score, assessing improvement on 6 levels.

Other studies used a different score in its entirety: Lee et al. [30] used the whole Oxford-12 score to define the MCID for the Knee Society rating system, taking advantage of the known Oxford-12 MCID of 5 points' improvement. Kuo et al [25] used a 100-point 4-question score for which there was consensus: e.g., that 25 points corresponds to severe dissatisfaction. Van der Wees et al. [31] used a different consensual breakdown of the Oxford-12 score (>41, excellent; 34-41, good; 27-33, moderate; <27 poor), defining the MCID as improvement by 1 category. Using a different score in its entirety requires interpretation to be known, which usually means an expert consensus. Whether one item or the whole score is used, the score

needs to be based on several response levels, to enable distinctions to be made in the population, like with anchors having between 3 and 15 possible responses [32]. Responses often need to be grouped according to the presence or lack of improvement.

4.3 Can an item serve as anchor for its own score?

Using an item to serve as an anchor for its own score seems reasonable but requires assessing in other studies as we were not able to find any other hip series using this method. Using an anchor is advantageous since it does not require the use of supplementary questions, and therefore confers a degree of objectivity as the patient does not realize that the assessment is being made. However, several anchors, or at least several methods, are often necessary [4].

4.4 Study limitations

The present study had several limitations;

1) The choice of the anchor was subjective; however, asking at least 2 questions enhances the strength of the MCID value.
2) Categorizing based on item results is debatable. The choice was supported by the number of patients per subgroup, which was close to the number for the reference anchor (1-5 point improvement). To calculate the MCID, responses often needed to be grouped. In a study of the MCID for the HOOS joint reconstruction score, Hung et al. [33] used a 7-question anchor, classifying patients from -3 to +3, then grouped patients reporting slight improvement (+1) together with slight deterioration and no change.
3) The sample size was small. However, Terwee et al. [34] considered 100 patients as sufficient to assess score properties. In a review of the literature, Copay et al. [5] advocated using the MDC95 to limit the impact of sample size.
4) The population was selected using strict criteria however, the MCID varies between populations. Our criteria limited bias, as scores often vary widely, as, for example, Marot et al. [35] showed for the KOOS, with variation according to age, gender and body-mass index.

5. Conclusion

The MCID in French-speaking participants with previous THR was comparable to other reported. Using a score item to determine the MCD appears to be possible if the item concerns general assessment.

Disclosure of interest: Sophie Putman is a consultant with Corin, but has no conflicts of interest to disclose in relation to the present study. Gilles Pasquier is an education and research consultant for Zimmer, but has no conflicts of interest to disclose in relation to the present study. Henri Migaud is Editor in Chief of Orthopaedics & Traumatology: Surgery & Research and, outside the scope of the present study, is an education and research consultant for Zimmer, Corin, MSD and SERF. Julien Girard is a consultant with Smith & Nephew, Corin and Mathys, but has no conflicts of interest to disclose in relation to the present study. The other authors have no conflicts of interest to disclose in relation to the present study or elsewhere.

Funding: None

Author contributions: Sophie Putman, Julien Dartus, Gilles Pasquier, Julien Girard and Henri Migaud collected the data and contributed to article writing; Alain Duhamel and Cristian Preda contributed to statistical analysis and article writing.

References

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control ClinTrials1989;10:407-15.

2. Singh JA, Schleck C, Harmsen S, Lewallen D. Clinically important improvement thresholds for Harris Hip Score and its ability to predict revision risk after primary total hip arthroplasty. BMC MusculoskeletDisord2016;17:256.

3. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin Orthop Relat Res 2018;476:2432-41.

4. 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;61:102-9.

5. Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum Clinically Important Difference: Current Trends in the Orthopaedic Literature, Part II: Lower Extremity: A Systematic Review. JBJS Rev2018;6:e2. doi: 10.2106/JBJS.RVW.17.00160.

6. Celik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. J Clin Epidemiol 2019;113:44-57.

7. Nilsdotter AK, Lohmander LS, KlässboM, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003;4:10.

8. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. Osteoarthritis Cartilage 2010;18:522-529.

9. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524-34.

10. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 1996;78:185-90.

11. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 Hip score to the French speaking population Orthop Traumatol Surg Res 2009;95:89-99.

12. Klouche S, Giesinger JM, Sariali EH. Translation, cross-cultural adaption and validation of the French version of the Forgotten Joint Score in total hip arthroplasty. Orthop Traumatol Surg Res 2018;104:657-661.

13. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7:541-6.

14. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582-592.

15. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat 2004;14:97-110.

16. Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd Ed. Chapter 5, John Wiley and Sons, New York, NY (2000), pp. 160-164

17. Youden WJ. Index for rating diagnostic test. Cancer 1950;3:32-5.

18. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Edition. Hillsdale: Lawrence Erlbaum, New York 1988.

19. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol 2015;68:73-9.

20. Martín-Fernández J, Gray-Laymón P, Molina-Siguero A, Martínez-Martín J, García-Maroto R, et al. Cross-cultural adaptation and validation of the Spanish version of the Oxford Hip Score in patients with hip osteoarthritis. BMC Musculoskelet Disord 2017;18:205. doi: 10.1186/s12891-017-1568-3.

21. Impellizzeri FM, Mannion AF, Naal FD, Hersche O, Leunig M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. Osteoarthritis Cartilage 2012;20:638-45

22. Nwachukwu BU, Chang B, Rotter B, Kelly BT, Ranawat AS, Nawabi DH. Minimal clinically important difference and substantial clinical benefit after revision hip arthroscopy. Arthroscopy 2018;34:1862-1868

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

24. Goodman SM, Mehta BY, Mandl LA, Szymonifka JD, Finik J, et al. Validation of the Hip Disability and Osteoarthritis Outcome Score and Knee Injury and Osteoarthritis Outcome Score pain and function subscales for use in total hip replacement and total knee replacement clinical trials. J Arthroplasty 2019;35: In Press. doi: 10.1016/j.arth.2019.12.038.

25. Kuo AC, Giori NJ, Bowe TR, Manfredi L, Lalani NF, Nordin DA, Harris AHS. Comparing methods to determine the Minimal Clinically Important Differences in Patient-Reported Outcome Measures for veterans undergoing elective total hip or knee arthroplasty in veterans health administration hospitals. JAMA Surg 2020;In Press. doi: 10.1001/jamasurg.2020.0024.

26. Giesinger JM, Behrend H, Hamilton DF, Kuster MS, Giesinger K. Normative values for the Forgotten Joint Score-12 for the US general population. J Arthroplasty 2019;34:650-655.

27. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670-5.

28. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. Arthritis Rheum 2009;61:313-20.

29. Danoff JR, Goel R, Sutton R, Maltenfort MG, Austin MS. How much pain is significant? Defining the Minimal Clinically Important Difference for the Visual Analog Scale for pain after total joint arthroplasty. J Arthroplasty 2018;33(Suppl7):S71-S75.

30. Lee WC, Kwan YH, Chong HC, Yeo SJ. The minimal clinically important difference for Knee Society Clinical Rating System after total knee arthroplasty for primary osteoarthritis. Knee Surg Sports Traumatol Arthrosc 2017;25:3354-3359.

31. Van der Wees PJ, Wammes JJ, Akkermans RP, Koetsenruijter J, Westert GP, et al. Patient-reported health outcomes after total hip and knee surgery in a Dutch University Hospital Setting: Results of twenty years clinical registry. BMC Musculoskelet Disord2017;18:97. doi: 10.1186/s12891-017-1455-y.

32. Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. Otolaryngol Head Neck Surg 2019;161:551-560.

33. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the Patient-Reported Outcomes

Measurement Information System Physical Function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics. World J Orthop 2018;9:41-49.

34. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651-7.

35. Marot V, Murgier J, Carrozzo A, Reina N, Monaco E, Chiron P, Berard E, Cavaignac E. Determination of normal KOOS and WOMAC values in a healthy population. Knee Surg Sports Traumatol Arthrosc 2019;27:541–548.

Figure legends

Figure 1: Definition of improvement on a HOOS item. Possible change ranges from -4 to +4. Results are then grouped to have a score from -2 to +2, where 0 represents no change, -1 and -2 improvement, and +1 and +2 worsening.

Figure 2: Definition of improvement on an Oxford-12 item. Possible change ranges from -4 to +4. Results are then grouped to have a score from -2 to +2, where 0 represents no change, -1 and -2 improvement, and +1 and +2 worsening.

Figure 3: ROC curves of anchors used for all scores. Y-axis: sensitivity; X-axis, 1- specificity. The 45° line corresponds to AUC 0.5 (non-discriminating).

Table 1: Distribution of pre- and post-operative scores. Results reported as mean [range] +/- standard deviation; * Δ: mean difference; **: pre- to post-operative p-value (matched Student test). On Cohen's criteria, effect size 0.2 is "weak", 0.5 "moderate" and 0.8 "strong".

	N	Preoperative	6-12 months' FU	Δ*	Effect size	p**
Oxford-12	123	19 [3-35] +/-8	40,8 [8-48] +/-10	21 [-12, +43] +/-11	1,92	<0.001
HOOS Symptom	123	41 [0-85] +/-17	83 [15-100] +/-19	42 [-25, +90] +/-24	1,76	<0.001
HOOS Pain	123	40 [2-75] +/-14	86 [17-100] +/-19	46 [-15, +85] +/-20	2,26	<0.001
HOOS Function	123	41 [3-88] +/-16	85 [19-100] +/-19	44 [-30, +88] +/-23	1,93	<0.001
HOOS Sport	123	23 [0-81] +/-17	75 [0-100] +/-25	51 [-18, +100] +/-26	1,92	<0.001
HOOS Quality of Life	123	24 [0-86] +/-18	78 [0-100] +/-24	54 [-50, +100] +/-28	1,93	<0.001
Forgotten joint score (FJS)	123		71 [0-100] +/-29			

Table 2: Area under the curve for various anchors with 95% confidence intervals. The greater the AUC, the more discriminating the anchor; AUC 0.5 is non-discriminating.

Area under the ROC curve	**Improvement anchor, 1-5 points**	**Oxford-12 question 1**	**HOOS question Q4 (quality of life)**	**Oxford-12 question 2**	**HOOS question S1 (symptom)**
Oxford-12	0.825 [0.763-0.907]	0.869 [0.806-0.932]	0.869 [0.806 0.932]	0.531 [0.402-0.661]	0.648 [0.546-0.720]
HOOS Symptom	0.819 [0.738-0.900]	0.839 [0.767-0.912]	0.827[0.736 .891]	0.427 [0.311-0.643]	0.693 [0.592-0.783]
HOOS Pain	0.761 [0.664-0.859]	0.877 [0.811-0.944]	0.837[0.751 .903]	0.479[0.357-0.602]	0.688 [0.580-0.796]
HOOS Function	0.788 [0.702-0.873]	0.830 [0.757-0.904]	0.813[0.762 .909]	0.514 [0.393-0.630]	0.631 [0.518-0.744]
HOOS Sport	0.808 [0.726-0.890]	0.879 [0.819-0.940]	0.850[0.782 .917]	0.420 [0.310-0.529]	0.606 [0.501-0.711]
HOOS Quality of life	0.822 [0.736-0.907]	0.859 [0.789-0.924]	0.920 [0.900 0.979]	0.497 [3.679-0.616]	0.645 [0.544-0.744]

Table 3: Distribution of scores according to anchor. N= number of subjects; ΔM = pre- to post-operative deference in mean. Blue= "worse of much worse"; Green = "slightly worse"; Yellow = "no change"; Orange = "slightly better"; Red = "better or much better"

Score	**Improvement anchor, 1-5**		**ΔM [range]+ /- standard deviation**	**Oxford question 1 anchor**		**ΔM [range]**	**HOOS question Q4 anchor**		**ΔM [range]**
		N			N			N	
Oxford-12	1	5	-3.8 [-13. +6] +/-8	-2	0		+2		
	2	0		-1	5	-5 [-14. +2] +/-6	+1	3	-1 [-12. +10] +/-15
	3	4	9 [-3. +10] +/-10	0	5	8 [0. +14] +/-6	0	6	1 [-10. +10] +/-7
	4	30	15 [-6. +36] +/-9	+1	40	15 [-6. +31] +/-8	-1	38	14[-6. +33]+/-8
	5	84	25 [7. +43] +/-8	+2	73	27 [9 . +43] +/-8	-2	76	26 [8. +43] +/-8
HOOS Symptom	1	5	-9 [-25. +5] +/-12	-2	0		+2		
	2	0		-1	5	-10 [-25. +0] +/-10	+1	3	-7.5 [-25. +10] +/-24
	3	4	13 [0. +40] +/-19	0	5	16 [5. 30] +/-11	0	6	0 [-15. +25] +/-20
	4	30	30 [-10. +70] +/-20	+1	40	31 [-10. +70] +/-20	-1	38	31 [0. +75] +/-14
	5	84	51 [5. 90] +/-19	+2	73	54 [10. 90] +/-14	-2	76	53 [10.90] +/-18
HOOS Pain	1	5	-2 [-15. +15] +/-14	-2	0		+2		
	2	0		-1	5	-5 [-55. +12] +/-12	+1	3	-3 [-15. +10] +/-17
	3	4	+15 [-15. +40] +/-22	0	5	24 [-3. +50] +/-23	0	6	4 [-12. +12] +/-20
	4	30	39 [3. +72] +/-20	+1	40	36 [+3. +72] +/-15	-1	38	38 [-4. +72] +/-17

	5	84	53 [25. +85] +/-14	+2	73	57 [32. +85] +/-12	-2	76	55 [25. +85] +/-13
HOOS Function	1	5	-4 [-19. +16] +/-16	-2	0		+2		
	2	0		-1	5	-9 [-19. 0] +/-8	+1	3	-6 [-16. +30] +/-15
	3	4	18 [-1. +32] +/-16	0	5	29 [5. +52] +/-15	0	6	2 [-20. 26] +/-15
	4	30	35 [0. +75] +/-19	+1	40	34 [-5. +75] +/-20	-1	38	34 [-1. +75] +/-18
	5	84	51 [30. +88] +/-18	+2	73	54 [18. +88] +/-15	-2	76	53 [30. 88] +/-18
HOOS Sport	1	5	-6 [-18. +12] +/-12	-2	0		+2		
	2	0		-1	5	-3 [-18. +18] +/-14	+1	3	0 [-19. +19] +/-26
	3	4	5 [0. +12] +/-6	0	5	18 [-12. +20] +/-30	0	6	9 [-12. +37] +/-20
	4	30	40 [0. +100] +/-21	+1	40	36 [-12 .81] +/-19	-1	38	36 [-12. +75] +/-22
	5	84	61 [0. +100] +/-20	+2	73	65 [6. +100] +/-19	-2	76	64 [6. +100] +/-19
HOOS Quality of life	1	5	-9 [-50. +3] +/-23	-2			+2		21[-12. +43] +/-11
	2	0		-1		-9 [-50. +25] +/-23	+1	3	-25 [-50. 0] +/-30
	3	4	29 [-10. +56] +/-23	0		28 [0. +56] +/-27	0	6	3 [-6. +18] +/-8
	4	30	39 [-6. +100] +/-27	+1		39 [-6. +93] +/-24	-1	38	35 [0. +69] +/-18
	5	84	64 [6. +100] +/-19	+2		68 [25. +100] +/-17	-2	76	70 [37. +100] +/-16

Table 4: MCID per score and per method. Mean MCID calculated per method, then for both methods.

	Distribution-based		Improvement anchor, 1-5	Oxford question 1 anchor	HOOS question Q4 anchor	Mean MCID for distribution-based method	Mean MCID for anchor-based method	Mean MCID for both methods
	SD_{Δ} *0,5	MDC95 %						
Oxford-12	5 ,5	7	11	12	12	6	12	9
HOOS Symptom	12	15	20	26	27	13	24	20
HOOS Pain	10	16	32	35	36	13	34	26
HOOS Function	11,5	12	29	30	28	11	29	22
HOOS Sport	13	22	31	32	34	17	32	26
HOOS Quality of life	14	17	26	28	26	15	27	22

Table 5: Oxford-12 hip MCID in the literature, according to country and method

	Number of patients	Country	Procedure	Method	MCID
Beard et al. [19]	82415	USA	Arthroplasty	Anchor	7.5
Fernandez et al. [20]	361	Spain	Arthroplasty	Anchor	7
Impelizari et al. [21]	102	Switzerland	Arthroscope	Anchor	8-10
Present series	123	France	Arthroplasty	Anchor and distribution	9 [range, 5.5-12]

Table 6: MCID for HOOS in the literature, according to country and method.

	Number of patients	Country	Procedure	Method	MCID
Nwachukwu et al. [22]	49	USA	Arthroscopy	Distribution	Sport 13 Function 7,9
Paulsen et al. [23]	1288	Denmark	Arthroplasty	Anchor and distribution	Pain 24 Quality of life 17
Lyman et al. [3]	2323	USA	Arthroplasty	Anchor	Pain 36 Symptoms 20 Function 14 Quality of life 13
Goodman et al. [24]	4801	USA	Arthroplasty	Anchor	Pain 22 Function 18
Kuo et al. [25]	271	USA	Arthroplasty	Anchor	Pain 25 Symptoms 20 Function 19 Quality of life 7 Sport 16
Present series	123	France	Arthroplasty	Anchor and distribution	Pain 26 [10-36] Symptoms 20 [12-27] Function 22 [11-28] Quality of life 23 [14-28] Sport 26 [13-34]

Q4. In general, how much difficulty do you have with your hip?

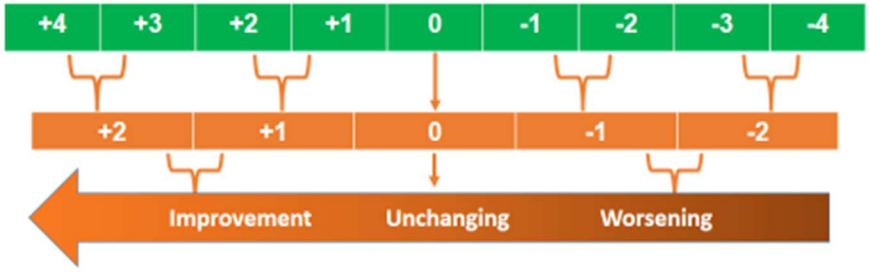
□ none	□ Mild	□ Moderate	□ severe	□ Extreme
0 point	1 point	2 points	3 points	4 points

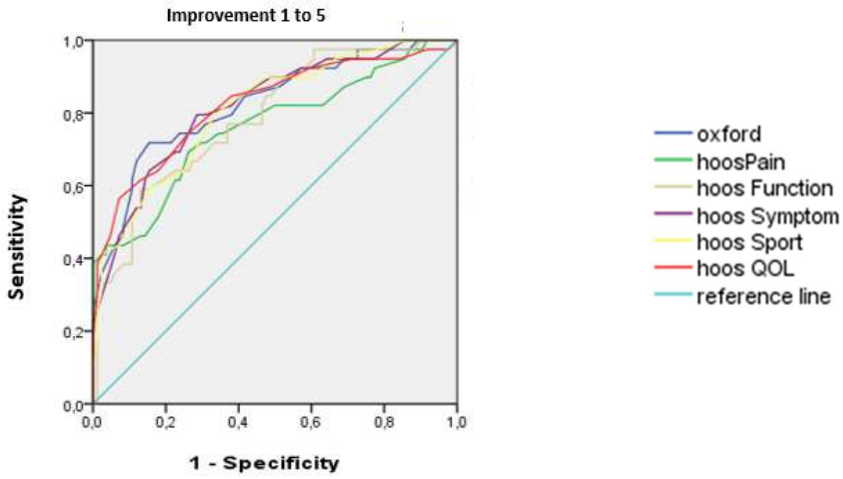
Possible difference between response before surgery and response after surgery

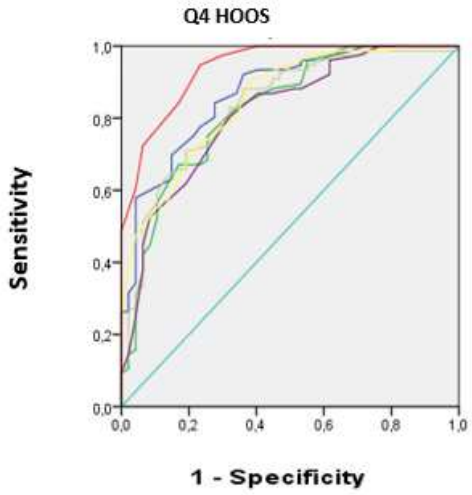
How would you describe the pain you usually have in your hip?

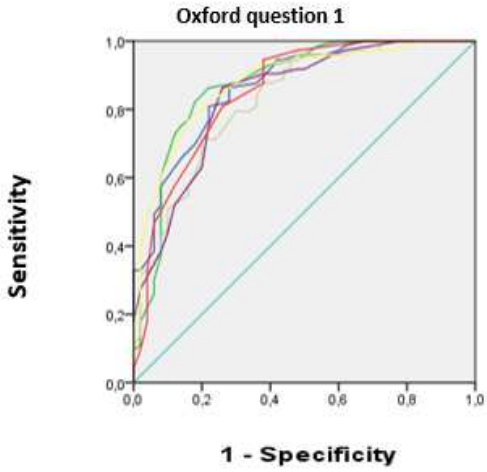
□ None	□ Very Mild	□ Mild	□ Moderate	□ Severe
4 points	3 points	2 points	1 point	0 points

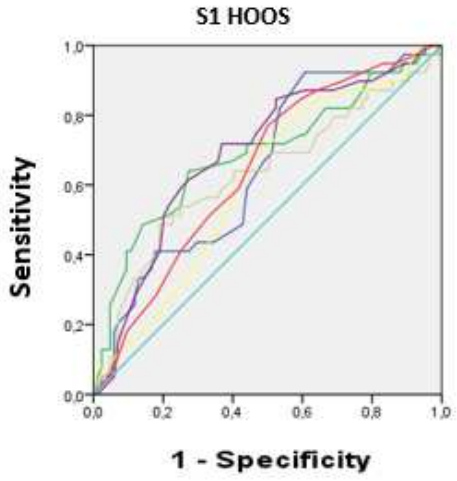
Possible difference between response before surgery and response after surgery

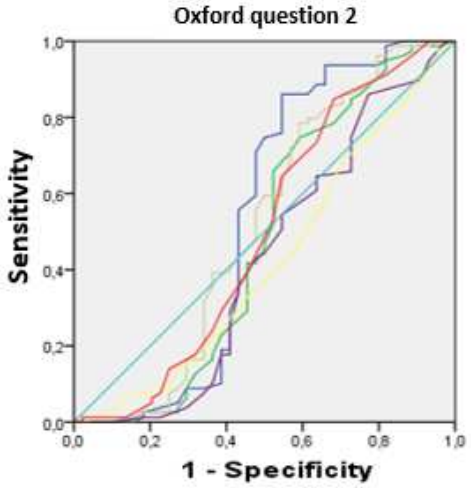



Improvement 1 to 5
Sensitivity
1 - Specificity
0,0
0,2
0,4
0,6
0,8
1,0
oxford
hoosPain
hoos Function
hoos Symptom
hoos Sport
hoos QOL
reference line


Q4 HOOS
Sensitivity
1 - Specificity
0,0
0,2
0,4
0,6
0,8
1,0


Oxford question 1
Sensitivity
1 - Specificity
0,0
0,2
0,4
0,6
0,8
1,0


S1 HOOS
Sensitivity
1 - Specificity
0,0
0,2
0,4
0,6
0,8
1,0


Oxford question 2
Sensitivity
1 - Specificity
0,0
0,2
0,4
0,6
0,8
1,0