Defining the Minimal Clinically Important Difference in Athletes Undergoing Arthroscopic Correction of Sports-Related Femoroacetabular Impingement

The Percentage of Possible Improvement

Patrick Carton,[*†‡] MD, FRCS, and David Filan,[†‡] MSc

Investigation performed at The Hip and Groin Clinic, UPMC Whitfield, Waterford, Ireland

Background: Measures of clinically meaningful improvement in patient-reported outcomes within orthopaedics are becoming a minimum requirement to establish the success of an intervention.

Purpose: To (1) define the minimal clinically important difference (MCID) at 2 years postoperatively in competitive athletes undergoing hip arthroscopic surgery for symptomatic, sports-related femoroacetabular impingement utilizing existing anchor- and distribution-based methods and (2) derive a measure of the MCID using the percentage of possible improvement (POPI) method and compare against existing techniques.

Study Design: Cohort study (diagnosis); Level of evidence, 2.

Methods: There were 2 objective outcome measures—the modified Harris Hip Score (mHHS) and 36-Item Short Form Health Survey (SF-36)—administered at baseline and 2 years postoperatively. External anchor questions were used to determine the MCID through mean change, mean difference, and receiver operating characteristic (ROC) techniques. Distribution-based calculations consisted of 0.5 SD, effect size, and standard error of measurement techniques. The POPI was calculated alongside each technique as an achieved percentage change of maximum available improvement for each athlete relative to the individual baseline score. The impact of the preoperative baseline score on the MCID was assessed by assigning athletes to groups determined by baseline percentiles. Statistical analysis was performed, with $P < .05$ considered significant.

Results: There were 576 athletes (96% male; mean age, 25.9 ± 5.7 years). The MCID score change (and POPI) for the mHHS and SF-36 ranged from 2.4 to 16.7 (21.6%-63.6%) and from 3.3 to 24.9 (22.1%-57.4%), respectively. The preoperative threshold value for achieving the ROC-determined MCID was 80.5 and 86.5 for the mHHS and 70.1 and 72.4 for the SF-36 for the patient-reported outcome measure (PROM) score– and POPI-calculated MCID, respectively. Through the commonly used mean change method, 40.0% (mHHS) and 42.4% (SF-36) of athletes were unable to achieve the MCID because of high baseline scores and PROM ceiling effects compared with 0% when the POPI technique was used. A highly significant difference for the overall MCID was observed between preoperative baseline percentile groups for the mHHS ($P = .014$) and SF-36 ($P = .004$) (improvement in points), while there was no significant difference between groups for either the mHHS ($P = .487$) or SF-36 ($P = .417$) using the POPI technique.

Conclusion: The MCID defined by an absolute value of improvement was unable to account for postoperative progress in a large proportion of higher functioning athletes. The POPI technique negated associated ceiling effects, was unrestricted by the baseline score, and may be more appropriate in quantifying clinically important improvement.

Keywords: MCID; percentage of possible improvement; hip arthroscopic surgery; athletes; femoroacetabular impingement

Hip arthroscopic surgery as a treatment intervention for symptomatic femoroacetabular impingement (FAI) is well established. Outcomes from this preservation surgery have been shown to significantly improve pain and physical function, particularly effective within the athletic population.[3,26,37,41] The change in the patient symptomatic state is evaluated with the use of evidence-based patient-reported outcome measures (PROMs); measurable differences will generally be evaluated using sensitive statistical tests to assess whether the change represents a true treatment difference as opposed to simply occurring by chance.[3,24]

The Orthopaedic Journal of Sports Medicine, 8(1), 2325967119894747
DOI: 10.1177/2325967119894747

While reports of statistical significance provide important insight and helpful information on the outcomes of specific interventions, this generalized reporting in isolation may be less informative on a case-by-case basis as well as run the risk of finding statistical relationships in the absence of clinical importance to patients and clinicians.[16,17,29,32]

The focus therefore on determining a benchmark by which change may be considered clinically meaningful for the individual patient may be more valuable. The concept of the minimal clinically important difference (MCID) was developed with this in mind and attempts to define the smallest change in a treatment outcome that a patient would identify as beneficial[16,17] and importantly may reflect the level of either an improvement or worsening in outcomes for the patient.[21]

The primary purpose of this study was to define the MCID at 2 years postoperatively in a cohort of competitive athletes undergoing hip arthroscopic surgery for symptomatic, sports-related FAI[5] utilizing existing anchor- and distribution-based methods for the MCID calculation. The secondary purpose of this study was to derive a measure of the MCID using the percentage of possible improvement (POPI) method and compare this with existing techniques.

We hypothesized that the change required and the optimal threshold needed to achieve the MCID within this athletic cohort as a whole are variable, depending on the method used and largely dictated by preoperative baseline scores; the assumption was that the MCID should be expected to be greater for those with lower baseline PROM scores. We also hypothesized that the POPI method may have benefits over existing techniques by negating the ceiling effect and minimizing the association with preoperative baseline scores.

METHODS

Analysis of prospectively collected data from our institutional hip preservation registry between January 2009 and October 2016 was undertaken. All participants provided signed consent for the use of their data to be included in our hip registry, which received institutional board approval. Athletes were included in this study if they underwent primary hip arthroscopic surgery for symptomatic, sports-related FAI and were competitively involved in sports (Gaelic football, hurling, soccer, rugby, and athletics) at the time of the initial presentation. Preoperatively, all athletes underwent a physical examination, dual-operator (handheld goniometric) evaluation of hip range of motion (ROM), and standardized radiographic assessment (anteroposterior [AP], false profile, and Dunn views) to quantify the degree of abnormal bony morphology.

Arthroscopic surgery in all cases was performed by a single experienced hip surgeon (P.C.) and consisted of femoroplasty and acetabuloplasty to correct FAI bony abnormalities at the femoral head-neck junction and acetabular rim, respectively. The acetabular labrum was preserved and repaired using a labral cuff repair technique, preserving the chondrolabral interface when possible[4]; in cases with significant chondrolabral separation, looped repair was utilized. Postoperatively, all athletes completed a standardized rehabilitation program over a 12-week period before returning to training. Standardized radiological views were obtained again at 6 weeks postoperatively to permit the measurement of bony deformity correction.

Internationally validated PROMs were utilized to assess joint-specific and generalized health status preoperatively and again 2 years postoperatively. PROMs specific for MCID analysis consisted of the modified Harris Hip Score (mHHS), one of the most commonly utilized PROMs to assess outcomes after hip arthroscopic surgery,[19] and the 36-Item Short Form Health Survey (SF-36), a health assessment tool focusing on the physical and mental aspects of general well-being.[15] The University of California, Los Angeles (UCLA) activity scale, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and a general satisfaction questionnaire were administered as additional outcome measures.

Exclusion criteria consisted of adverse preoperative radiographic features including Tönnis grade >1, Perthes disease, dysplasia (defined by lateral center-edge angle [LCEA] <20°), >40 years of age, and cases undergoing revision surgery within 2 years.

MCID Calculation

MCID values were calculated using 3 anchor-based and 3 distribution-based techniques. For the anchor-based approach, 3 domain-specific questions and 1 general satisfaction question were asked at 2 years after hip arthroscopic surgery: (1) Pain (considered to be one of the most debilitating factors,[13] particularly for athletes): "How well did the surgery on your joint relieve your pain?" (2) Abilities: "How well did the surgery on your joint increase your ability to perform regular activities?" (3) Sports: "How well did the surgery on your hip joint allow you to perform sports activities?" and (4) Expectations: "How well did the surgery on your joint meet your expectations?" Choice responses included "excellent," "very good," "good," "fair," and "poor."

We considered a response of "fair" to equate to an MCID from baseline to 2 years postoperatively, and a "poor"

*Address correspondence to Patrick Carton, MD, FRCS, The Hip and Groin Clinic, UPMC Whitfield, Cork Road, Suite 5, Butlerstown North, Waterford, Ireland (email: cartoni2k@hotmail.com).

†The Hip and Groin Clinic, UPMC Whitfield, Waterford, Ireland.

‡UPMC Whitfield, Waterford, Ireland.

Final revision submitted September 17, 2019; accepted September 20, 2019.

The authors declared that there are no conflicts of interest in the authorship and publication of this contribution. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval was not sought for the present study.

response indicated a failure to meet this level of improvement. Patients who answered "fair" and demonstrated a detectable increase in PROM scores were assigned to group 1 (improved); all patients answering "poor," regardless of changes in PROM scores, were assigned to group 2 (nonimproved).

The MCID according to the anchor-based approach was calculated using 3 methods:

1. The mean change[16] in PROM scores for those athletes in the improved group.
2. The mean difference[33] in PROM score improvements between those who improved (group 1) versus those who did not improve (group 2).
3. A receiver operating characteristic (ROC) curve to discriminate between improved (group 1) and nonimproved (group 2) athletic satisfaction; an area under the curve (AUC) >0.7 was considered to be acceptable and >0.8 considered excellent.[8]

The MCID according to the distribution-based approach was calculated using 3 methods:

1. A 0.5 SD, calculated by obtaining half the SD of the measured change in PROM scores.[27]
2. The effect size, calculated by multiplying the SD of the baseline score by 0.2 (small effect size).[36]
3. The standard error of measurement, calculated by multiplying the SD of the baseline score by the square root of 1 minus the intraclass correlation coefficient.[7]

Percentage of Possible Change

The percentage of possible change was determined relative to the direction of PROM scores at 2 years; that is, if athletes' PROM scores increased, then the improvement was calculated as a percentage relative to the maximum achievable improvement for that athlete (POPI). Similarly, if an athlete's PROM scores decreased, this decrease from baseline was calculated as a percentage relative to the maximum possible decrease available to that athlete (POPD). For the mHHS and SF-36, the maximum possible score was 100; the lowest recorded scores for the mHHS (44) and SF-36 (23) in this cohort of athletes were used as the maximal lowest score to improve the accuracy of calculations by avoiding underestimating the POPD value (avoiding an unrealistic maximal score of zero for either outcome test).

All cases of improved PROM scores at 2 years were calculated as the following: *POPI = (PROM change / (maximum possible achievable score – preoperative score))* × *100*. All cases of decreased PROM scores at 2 years were calculated as the following: *POPD = (PROM change / (preoperative score – lowest PROM score))* × *100*. The MCID was subsequently calculated using anchor-based methods (described above).

Impact of Preoperative Baseline Scores on MCID

Athletes who had "fair" satisfaction and an increase in PROM scores (group 1, improved) were divided into 4 different subgroups (A-D) for the mHHS and SF-36 (according to the baseline percentiles). The MCID was calculated using the mean change method (most commonly utilized method for deriving the MCID[1]) for PROM scores and the POPI from baseline to 2 years postoperatively for each of these subgroups (and specific to each of the 4 anchor-based questions) and analyzed to determine whether an association was evident between initial baseline scores and the calculated value of the MCID. Additionally, the ability for athletes from the entire cohort to achieve a calculated MCID was explored, specifically relative to a baseline threshold PROM score.

Statistical Analysis

The independent-samples t test and chi-square test were utilized to test for differences between groups for the MCID calculation. Nonparametric analysis (median with interquartile range) was performed to examine between-group (Mann-Whitney U test) and within-group (Wilcoxon signed-rank test) differences both preoperatively and at 2 years postoperatively. Analysis of variance assessed the relationship of the mean change in PROM scores and the POPI when baseline PROM scores were categorized into percentiles. Threshold analysis was performed using nonparametric ROC analysis with the AUC to identify the baseline PROM score predictive of achieving the MCID. Bivariate analysis was conducted on both outcome measures to determine any factors associated with achieving the MCID for the main representative anchor question (Expectations). Independent variables measured on a continuous scale included age, LCEA, alpha angle (AA), degree of bony resection, and ROM improvement, while categorical measured variables included sex, presence of radiographically measured cam and pincer deformities, labral management (repair or no repair), Tönnis grade, and presence of the crossover sign. Multivariate forward stepwise logistic regression was then performed on all significant variables highlighted through the preliminary bivariate analysis. The baseline PROM score was adjusted for in the multivariable model. All statistical analyses were performed using SPSS Version 25.0 software (IBM); $P < .05$ was considered significant.

RESULTS

Demographics

A total of 576 athletes were included in this study (96% male). The mean age was 25.9 ± 5.7 years. At the time of surgery, all athletes were regularly involved in competitive sports training and/or competing <3 (15.6%), 3 to 5 (71.3%), and >5 days per week (13.1%), identifying Gaelic Athletic Association (GAA) hurling (46.9%), GAA football (36.5%), soccer (9.7%), rugby (4.9%), and athletics (2.1%) as their main competitive sport. Athletes were reviewed at a mean of 2.4 ± 0.7 years postoperatively (median, 2.1 years [interquartile range, 2.0-2.5 years]).

TABLE 1
Preoperative Patient Demographics and Characteristics[a]

	Improved (n = 43)	Nonimproved (n = 38)	*P*
Age, y	26.0 ± 6.4	26.3 ± 5.4	.835
Sex, n			.284
Male	43	37	
Female	0	1	
Lateral center-edge angle, deg	34.6 ± 6.6	33.1 ± 4.7	.263
Alpha angle, deg			
AP view	76.7 ± 114.6	73.8 ± 14.7	.383
Dunn view	63.6 ± 13.1	63.6 ± 14.8	.995
Tönnis grade, %			.519
0	64.3	71.1	
1	35.7	28.9	
Range of motion, deg			
Flexion	110.3 ± 11.4	112.8 ± 11.0	.307
Abduction	45.3 ± 8.2	45.4 ± 6.3	.952
Adduction	18.8 ± 8.8	20.0 ± 6.7	.502
External rotation	37.7 ± 8.7	37.1 ± 8.1	.733
Internal rotation	22.2 ± 11.3	25.4 ± 11.9	.219
mHHS score, median (IQR)	74 (70-83)	79 (73-86)	.065
SF-36 score, median (IQR)	66.5 (57.6-73.6) [n = 36]	69.9 (60.6-78.7) [n = 35]	.204

[a]Responses from the overall Expectations anchor question ("How well did the surgery on your joint meet your expectations?") were used to define the groups. Data are presented as mean ± SD unless otherwise indicated. AP, anteroposterior; IQR, interquartile range; mHHS, modified Harris Hip Score; SF-36, 36-Item Short Form Health Survey.

TABLE 2
Clinical Outcomes[a]

	Preoperative	2 y Postoperatively	*P*
Range of motion, deg			
Flexion	111.0 ± 11.2	117.5 ± 8.9	<.001
Abduction	44.8 ± 9.0	48.8 ± 8.7	<.001
Adduction	20.3 ± 7.8	24.3 ± 6.1	<.001
External rotation	37.6 ± 8.3	40.3 ± 7.5	<.001
Internal rotation	23.5 ± 10.9	31.2 ± 9.2	<.001
Total	237.2 ± 31.7	262.1 ± 27.8	<.001
PROM score, median (IQR)			
mHHS (n = 576)	82 (73-93)	96 (96-100)	<.001
SF-36 (n = 509)	75.4 (61.6-86.6)	90.9 (82.8-95.0)	<.001
UCLA activity scale (n = 576)	8 (6-10)	10 (9-10)	<.001
WOMAC (n = 493)	15 (6-28)	2 (0-8)	<.001
Radiographic measurement, deg			
Lateral center-edge angle	34.0 ± 6.1	30.4 ± 5.7	<.001
Alpha angle on Dunn view	59.8 ± 12.9	50.9 ± 10.0	<.001
Alpha angle on AP view	68.4 ± 17.5	61.4 ± 15.1	<.001

[a]Data are presented as mean ± SD unless otherwise indicated. AP, anteroposterior; IQR, interquartile range; mHHS, modified Harris Hip Score; PROM, patient-reported outcome measure; SF-36, 36-Item Short Form Health Survey; UCLA, University of California, Los Angeles; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

There was no statistical difference between the proportion of athletes answering "excellent" (P = .855), "very good" (P = .160), "good" (P = .546), "fair" (P = .251), or "poor" (P = .128) for any of the 4 anchor questions. As such, the Expectations anchor was considered representative of the overall level of satisfaction (Table 1); there was no statistical difference in baseline demographics between the improved ("fair" satisfaction) and nonimproved ("poor" satisfaction) groups used in the calculation of the anchor-based MCID (Table 1).

Radiographic and Hip ROM Evaluation

For all athletes, a localized acetabular rim deformity was observed on the false profile view; 68.8% had an LCEA >30° (37% with >35°),[5] and in 76.3% of athletes, a crossover sign (standardized AP view) was observed. In 66.6% of athletes, a cam deformity was present, observed in 54.7% on the Dunn view (AA >55°) and in 57.8% on the AP view (AA >65°); 77.9% were graded as Tönnis 0 and 22.1% as Tönnis 1. After surgery, the LCEA improved from 34.0° ± 6.1° to 30.4° ± 5.7° (n = 523; P < .001), the AA (Dunn view) from 59.8° ± 12.9° to 50.9° ± 10.0° (n = 408; P < .001), and the AA (AP view) from 68.4° ± 17.5° to 61.4° ± 15.1° (n = 530; P < .001) (Table 2).

For all athletes with available preoperative and postoperative ROM assessments (n = 410), there was a highly significant and clinically important improvement in hip motion (P < .001) (Table 2).

PROMs and Overall Satisfaction

There was a highly statistically significant improvement in all outcome scores (P < .001) (Table 2). Overall satisfaction with the surgery at 2 years postoperatively was calculated by determining the mean satisfaction from our 4 anchor questions. Overall satisfaction was described as "excellent" (35.6%), "very good" (32.3%), "good" (16.3%), "fair" (10.6%), and "poor" (5.5%), with 94.5% of athletes indicating that the surgery was beneficial.

MCID Calculation

The MCID for the mHHS ranged from 2.4 to 16.7 using the PROM score change and from 21.6% to 63.6% using the POPI. The percentage of all athletes who improved by the MCID ranged from 72.4% to 86.3% (PROM) and from 72.2% to 82.8% (POPI) (Tables 3 and 4).

The MCID for the SF-36 ranged from 3.3 to 24.9 using the PROM score change and from 22.1% to 57.4% using

TABLE 3
Anchor-Based Method of Calculating MCID Values[a]

	Pain (n = 33 Fair; n = 22 Poor)		Expectations (n = 43 Fair; n = 38 Poor)		Sports (n = 37 Fair; n = 38 Poor)		Abilities (n = 33 Fair; n = 29 Poor)	
mHHS	PROM	POPI, %	PROM	POPI, %	PROM	POPI, %	PROM	POPI, %
Mean change (fair)	14.9 (75.4)	63.6 (72.2)	15.2 (73.5)	62.6 (72.4)	16.0 (73.5)	60.5 (72.6)	13.5 (74.0)	59.4 (72.8)
Mean difference (fair – poor)	16.7 (72.4)	58.3 (72.9)	12.8 (78.4)	44.9 (78.0)	14.4 (75.4)	42.7 (79.0)	13.6 (74.0)	49.6 (77.3)
ROC curve (fair vs poor)	7.5 (86.3)	38.0 (80.3)	12.5 (78.4)	48.0 (77.3)	11.5 (81.3)	36.9 (80.4)	11.5 (81.3)	36.9 (80.4)
	Pain (n = 25 Fair; n = 19 Poor)		Expectations (n = 36 Fair; n = 35 Poor)		Sports (n = 32 Fair; n = 35 Poor)		Abilities (n = 29 Fair; n = 26 Poor)	
SF-36	PROM	POPI, %	PROM	POPI, %	PROM	POPI, %	PROM	POPI, %
Mean change (fair)	19.4 (55.9)	51.8 (54.0)	20.7 (58.6)	57.4 (49.9)	18.9 (55.2)	48.9 (55.4)	17.9 (56.1)	51.4 (54.0)
Mean difference (fair – poor)	24.9 (56.9)	51.7 (54.0)	19.7 (56.5)	57.4 (49.9)	21.8 (58.1)	47.8 (56.8)	21.0 (59.8)	52.6 (52.8)
ROC curve (fair vs poor)	6.9 (66.7)	24.0 (70.9)	6.9 (66.7)	28.3 (70.0)	6.5 (67.2)	28.3 (70.0)	6.0 (67.9)	28.3 (70.0)

[a]Data in parentheses indicate the percentage of athletes meeting the MCID (for all those athletes in whom it was possible to achieve the MCID based on baseline scores). MCID, minimal clinically important difference; mHHS, modified Harris Hip Score; POPI, percentage of possible improvement; PROM, patient-reported outcome measure; ROC, receiver operating characteristic; SF-36, 36-Item Short Form Health Survey.

TABLE 4
Distribution-Based Method of Calculating MCID Values[a]

	PROM	POPI, %
mHHS		
Effect size	2.4 (83.2)	N/A
0.5 SD	6.8 (86.3)	21.6 (82.8)
SEM	3.6 (82.3)	N/A
SF-36		
Effect size	3.3 (74.9)	N/A
0.5 SD	9.8 (64.3)	22.1 (71.8)
SEM	5.2 (68.6)	N/A

[a]Data in parentheses indicate the percentage of athletes meeting the MCID (for all those athletes in whom it was possible to achieve the MCID based on baseline scores). MCID, minimal clinically important difference; mHHS, modified Harris Hip Score; N/A, not available; POPI, percentage of possible improvement; PROM, patient-reported outcome measure; SEM, standard error of the mean; SF-36, 36-Item Short Form Health Survey.

the POPI. The percentage of athletes who improved by the MCID ranged from 55.2% to 74.9% (PROM) and from 49.9% to 71.8% (POPI) (Tables 3 and 4).

Impact of Preoperative Baseline Scores on MCID

A highly significant difference in the MCID was observed for each of the subgroups (A-D) for the mHHS other than for the Pain anchor (Expectations, $P = .014$; Sports, $P = .018$; Abilities, $P = .038$; Pain, $P = .147$) and for the SF-36 (Expectations, $P = .004$; Sports, $P = .000$; Abilities, $P = .002$; Pain, $P = .014$), regardless of the type of anchor-based question used. When the POPI was examined, however, there was no significant difference in scores between groups for either the mHHS (Pain, $P = .626$; Expectations, $P = .487$; Sports, $P = .898$; Abilities, $P = .291$) or SF-36 (Pain, $P = .512$; Expectations, $P = .417$; Sports, $P = .148$; Abilities, $P = .163$) (Table 5).

MCID Achievability

The improvement in PROM scores and POPI required to achieve the MCID was 15.2 points and 62.6% and 20.7 points and 57.4% for the mHHS and SF-36, respectively (Table 3). A total of 29 athletes (5.0%) had a maximal preoperative mHHS score and 2 athletes (0.4%) had a maximal preoperative SF-36 score of 100 points and therefore were unable to achieve the MCID because of the ceiling effect of the PROMs. In 219 athletes (40.0%), the preoperative score was too high for athletes to improve by the MCID for the mHHS and 215 athletes (42.4%) for the SF-36. There were no athletes unable to potentially achieve the MCID when using the POPI method.

Threshold MCID Calculation

The baseline score threshold predictive of achieving the MCID for the mHHS was 80.5 (AUC, 0.937 [95% CI, 0.918-0.956]) for the PROM score change and 86.5 (AUC, 0.694 [95% CI, 0.645-0.744]) for the POPI. For the SF-36, the baseline threshold predictive of achieving the MCID was 70.1 (AUC, 0.949 [95% CI, 0.932-0.966]) for the PROM score change and 72.4 (AUC, 0.733 [95% CI, 0.690-0.777]) for the POPI. Preoperative scores below these threshold values were predictive of achieving the MCID compared with scores above this threshold (Figure 1).

Multivariate Regression Analysis

There were no predictive variables for achieving the PROM score–calculated MCID for either the mHHS or

TABLE 5
Impact of Preoperative Baseline Scores on MCID Using Mean Change Method[a]

	mHHS					SF-36				
Percentile Subgroup	n	PROM Score Change	*P* Value	POPI, %	*P* Value	n	PROM Score Change	*P* Value	POPI, %	*P* Value
Pain			.147		.626			.014		.512
A	9	10.7		73.8		6	7.2		36.8	
B	8	13.4		63.2		7	17.2		58.0	
C	9	15.8		58.2		7	25.4		58.0	
D	7	21.0		56.7		5	29.0		52.3	
Expectations			.014		.487			.004		.417
A	12	9.8		70.6		8	9.1		44.8	
B	10	12.8		57.9		10	18.5		63.3	
C	12	15.7		54.9		10	22.9		60.0	
D	9	24.7		67.4		8	32.2		59.3	
Sports			.018		.898			.000		.148
A	9	8.9		62.6		8	7.4		32.8	
B	11	14.9		62.3		9	18.9		59.3	
C	8	18.8		62.7		7	19.9		49.1	
D	9	22.0		54.5		8	29.5		53.2	
Abilities			.038		.291			.002		.163
A	10	8.0		60.7		7	6.9		35.1	
B	7	15.1		70.0		7	17.0		57.7	
C	9	16.1		59.5		9	24.6		64.4	
D	7	16.3		42.8		6	21.9		43.3	

[a]Subgroups: A, >75th percentile; B, >50th-75th percentile; C, 25th-50th percentile; D, <25th percentile. MCID, minimal clinically important difference; mHHS, modified Harris Hip Score; POPI, percentage of possible improvement; PROM, patient-reported outcome measure; SF-36, 36-Item Short Form Health Survey.

SF-36. Multivariable models for predicting the POPI-calculated MCID were applied: For the mHHS, a larger increase in the total ROM change was predictive of achieving the MCID (odds ratio [OR], 1.016 [95% CI, 1.008-1.024]; $P < .001$), while a larger preoperative AA on the AP view was predictive of not achieving the MCID (OR, 0.985 [95% CI, 0.971-0.999]; $P = .024$). For the SF-36, female sex (OR, 0.121 [95% CI, 0.026-0.568]; $P = .007$) and a larger abduction change (OR, 1.036 [95% CI, 1.015-1.058]; $P = .001$) were significant clinical predictors of achieving the MCID.

DISCUSSION

This study consisted of young competitive athletes who underwent arthroscopic correction for symptomatic, sports-related FAI. Arthroscopic FAI correction resulted in highly statistically significant improvements in all PROMs 2 years postoperatively, with 94.5% indicating satisfaction with the outcome of the surgery.

The importance of adequate bony deformity correction and labral repair to achieving postoperative success has been previously documented.[20,22,34] In this cohort of competitive athletes, establishing normal radiological parameters and increasing impingement-free hip ROM were important goals to achieving optimal outcomes. The preservation and repair of essential soft tissue structures including the chondrolabral junction,[4] the labrum,[25] and the capsule[10] are central to the success of the procedure.

The objective surgeon-reported outcome measures and PROMs utilized in this study are indicative of a successful outcome at 2 years after hip arthroscopic surgery, but when presented in isolation, fail to demonstrate to what degree improvement is needed for patients to have true satisfaction with their outcome. The MCID is one measure of an important change in a patient's symptomatic state and, by its definition, corresponds to the minimum outcome change that a patient may consider meaningful.[16,21,23] In theory, the MCID may be a powerful metric, but in reality, much confusion exists surrounding the best method of its calculation and the true application of this outcome measure in clinical practice.

The mHHS has been previously used to calculate the MCID in patients undergoing arthroscopic FAI surgery.[6,18,30,39] The SF-36 has demonstrated good correlation with the mHHS, particularly the physical component subscale, and has been recommended for use in hip arthroscopic surgery populations in conjunction with a joint-specific instrument such as the mHHS.[31] Benchmarks for the MCID specific to the SF-36 are less well described, particularly within the arthroscopic literature.

Several methods of determining the MCID have been proposed in the literature to date[7,42] and are broadly assigned to distribution- or anchor-based models. MCID values for the mHHS have been reported ranging from 6.5 to 20.0 at 1 year after hip arthroscopic surgery.[6,18,30,39] The MCID values within our athletic cohort were similarly varied and dependent on the calculation method employed, despite all analyses performed on the single focused population. This limitation in the MCID has previously been reported Terwee et al[38] reported 5 MCID values for the

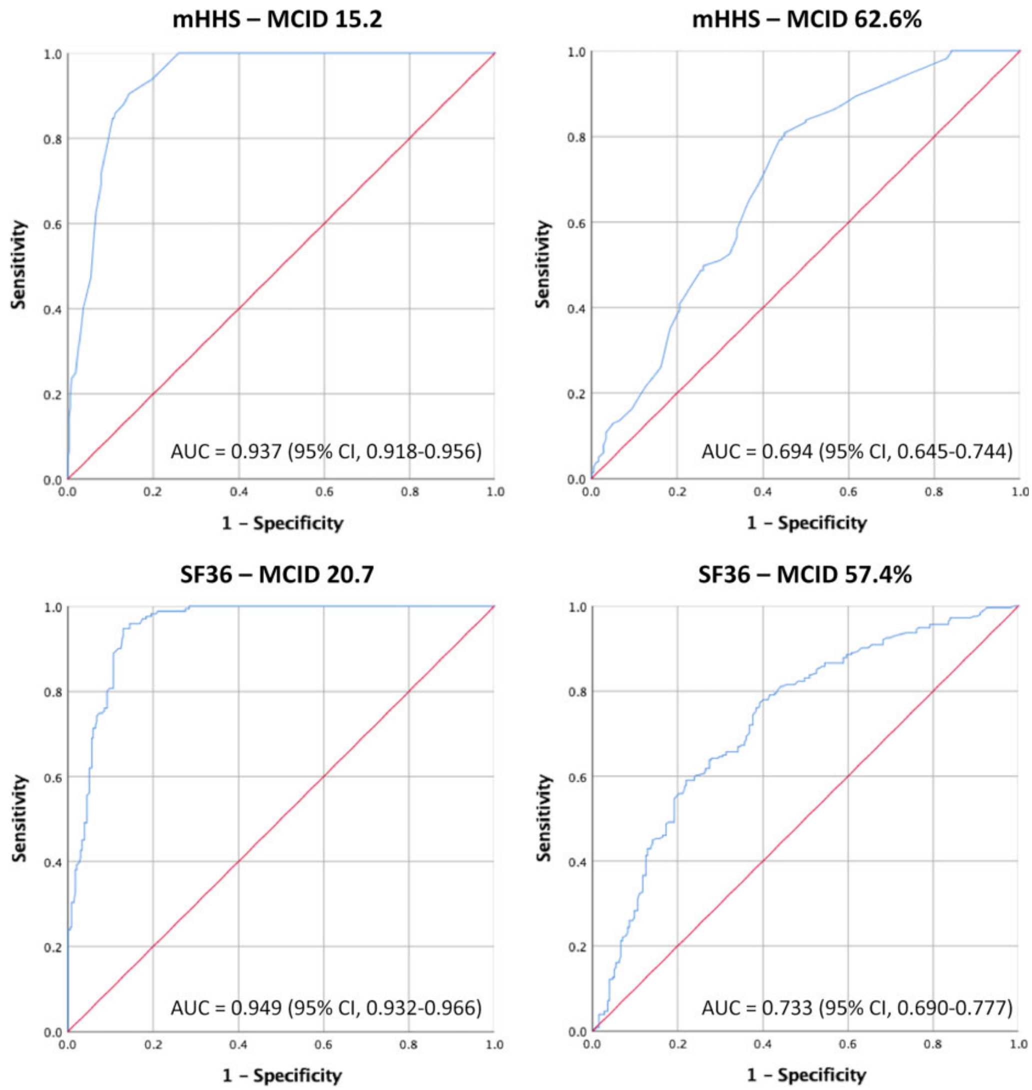

Figure 1. Receiver operating characteristic (ROC) curve with area under the curve (AUC) for the baseline patient-reported outcome measure (PROM) score threshold predictive of achieving the minimal clinically important difference (MCID). Diagonal segments are produced by ties. SF-36, 36-Item Short Form Health Survey.

physical function subscale of the WOMAC when using 5 different MCID methods on the same study population. Clinically significant outcome improvement may continue beyond 2 years postoperatively,[28] and as such, establishing MCID standards at this time point may be more valuable.

Distribution-based methods of determining a level of clinically associated improvement have been extensively used,[27,30,32,39] and the validity of their use depends on high test-retest and interrater reliability as well as internal consistency of the outcome measure being used.[14,18,30] There were 3 previously described methods for the calculation of the distribution-based MCID included in this study; there was considerable variation between the scores from each of these methods and also when compared with anchor-based MCID values. A further limitation of utilizing distribution-based methods is the inability to derive the level of patient benefit from the scores, as they are not compared with any measure of meaningful improvement.

For the anchor-based model, we utilized 4 anchor questions (3 domain-specific and 1 general) to establish the MCID; given our cohort of athletes, it was important to include an anchor question based on abilities, sports, pain, and overall expectation from the surgery. Pain and physical function anchor questions, particularly in a younger active population, have previously been shown to be useful for psychometric analysis of clinically meaningful outcome improvement.[29]

The number of divisions in the anchor response is arbitrary, but the more response levels included, the smaller the number of patients in each level and the greater the overlap of outcome scores between adjacent levels with less discriminatory value; to minimize this, we included 5 response levels: excellent, very good, good, fair, and poor.

For the MCID to function as a valuable metric, there needs to be adequate discrimination between those who have had meaningful improvement and those who have not. Using ROC curve analysis, our inclusion criteria yielded an acceptable strength of association (AUC >0.7) for the mHHS and excellent strength (AUC >0.8) for the SF-36. However, the arbitrary selection of the number of groups and divisions within the global rating scale and the inclusion criteria of these groups will result in significant variability of the MCID using ROC curve analysis and present major limitations to its use for calculating the MCID in general.

The mean difference method calculates the MCID as the difference between the baseline and postoperative scores for the athletes in the "fair" versus "poor" groups and as a result produced some of the highest MCID values. However, this method will not provide an MCID value that equates to meaningful improvement but a measure of the PROM score difference between those who have improved and those who have not and therefore does not reflect a true measure of the minimal clinically important change.

The mean change method calculates the MCID as the mean change in scores from baseline to postoperatively of the improved group (ie, those with "fair" satisfaction and an improvement in PROM scores); this value signifies a true change from baseline in only those categorized as having a minimal meaningful improvement and as such would seem to be representative of a true MCID.

The anchor-based MCID in this cohort of athletes ranged from 7.5 to 16.7 for the mHHS and from 6.0 to 24.9 for the SF-36. The previously reported MCID using the mean change method for the mHHS was 8,[18] and a net change of 11.31 was reported as slightly improved by Nwachukwu et al.[29] When we considered the mean change in PROM scores in the current study to represent the MCID, we found a change of 15.2 for the mHHS and 20.7 for the SF-36. Given that our cohort consisted of all competitive athletes, a higher value for the MCID might have been expected; however, as the preoperative baseline PROM scores were generally much higher than in similar nonathletic cohorts,[3,9,28,30,39] the scope for the increase in PROM scores was also lower.

Levy et al[23] performed a meta-analysis to examine the percentage of study populations that met the MCID for the mHHS. Using the MCID value of 8,[18] they observed that 97% of the study populations had a mean change from preoperative to postoperative scores greater than the MCID of 8. In Levy et al's meta-analysis of publications, primarily from the United States, the mean preoperative mHHS score was 61 ± 9.3, increasing to 83 ± 8.2 (a mean increase of 22 points). In our athletic cohort, the mean preoperative baseline score for the mHHS was 81 ± 12.0; despite increasing to a much higher postoperative score of 95 ± 8.1, this represents a smaller mean increase of 14. The comparison of the meta-analysis[23] and our study highlights the difficulty in using an MCID value derived from one particular method in different study populations and generalizing across multiple study cohorts. The higher the mean preoperative baseline PROM score, the lower the possible mean change and less likely to be able to meet the MCID.

Furthermore, even within a similar demographic as represented in this athletic cohort, the MCID as a measure of mean change is highly variable between subgroups relative to the baseline score. We were able to show this in the current study through significant differences across all 4 anchor questions for the SF-36 and 3 of 4 anchor questions (Expectations, Sports, and Abilities) for the mHHS. The MCID ranged from 10.7 to 21.0, with those subgroups with the lowest mean preoperative scores having the highest mean change (MCID) postoperatively and those with the highest mean preoperative scores having the lowest mean change (MCID). This clearly demonstrates that the MCID using the mean change method (and therefore also the mean difference method) is highly dependent on preoperative baseline scores.

The concept of utilizing the percentage change as a measure of meaningful improvement has been previously described: Farrar et al[11] reported a 30% change to represent a clinically important change (raw change/baseline × 100); Salaffi et al[35] considered a 15% reduction on the Numerical Rating Scale (pain scale) to represent the MCID; Tubach et al[40] reported a 20% relative improvement to represent the minimal clinically important improvement (MCII); and more recently, Bellamy et al[2] also looked at the relative change from baseline in patients with osteoarthritis, and they explained the importance of considering both the raw score change and relative score change. However, in all of these studies using a percentage relative change, the MCID is still dependent on the baseline scores and limited by the PROM ceiling effects.

The POPI considers how much patients have actually improved as a percentage of the maximum scope for improvement available to them relative to their preoperative symptomatic state, objectively measured with the use of validated PROMs. Gilmer et al,[12] in their study, used a similar concept of the percentage of total possible improvement to identify patients achieving a preselected MCID threshold value of 30%. In this study, however, the POPI and POPD were utilized specifically to calculate an actual MCID value; the MCID ranged from 21.6% to 63.6% for the mHHS and from 22.1% to 57.4% for the SF-36. The POPI method removes the inaccuracy associated with a mean improvement score or threshold particularly when trying to associate a subjective clinically meaningful satisfaction rating to an objective postoperative outcome measurement tool. The MCID represented through mean change as a POPI was not significantly different between any of the subgroups determined through preoperative baseline scores for either the mHHS or the SF-36. This demonstrated that the POPI method of calculating the MCID was independent of preoperative baseline scores.

The baseline and postoperative scores for the mHHS in this study are higher than in many similar studies, indicating that athletes may have a higher baseline score when compared with a more general population. The ability to achieve the MCID may be limited by a high preoperative baseline PROM score. In this study, 40.0% of athletes were unable to achieve the MCID for the mHHS and 42.4% of athletes were unable to achieve the MCID for the SF-36 because of a combination of high preoperative baseline scores and ceiling effects of PROMs in which a traditional method of the MCID calculation (mean score change) was utilized. The POPI introduced in the current study represents a more versatile method of determining the MCID, and by comparison, there were no athletes (0%) unable to potentially achieve the MCID utilizing this method, allowing for a more accurate consideration of overall success in which all surgical patients can be considered and followed up.

Limitations

A large number of competitive athletes were included in the overall cohort to allow for sufficient numbers in the "fair" improved group to permit the accurate calculation of the MCID; this required an extended study period for inclusion. Over this period, subtle changes in surgical techniques have evolved including preservation of the chondrolabral interface, labral cuff repair technique, and progressive introduction of capsular repair; these changes may have contributed to changes in clinical outcomes and the MCID, which were not specifically evaluated. In this cohort of competitive athletes, the mean preoperative baseline outcome scores were higher than those in other published studies, and expectations from surgery may be higher; as such, the results of this study may not be representative of a more generalized surgical population.

CONCLUSION

The anchor-based mean change method had the most relevance to the MCID. Serious deficiencies were evident, however, when the MCID was represented by a change in absolute scores because of the ceiling effect of PROMs and the dependence on baseline scores. In athletic cohorts in whom preoperative baseline scores are often high, the limitations of this method may be particularly apparent. This is overcome by using the POPI method, which negates any ceiling effect and is independent of baseline scores. The mean change method utilizing the POPI is recommended for measuring the MCID in a cohort of athletes undergoing arthroscopic hip surgery for symptomatic, sports-related FAI. This method is also suitable for other nonathletic cohorts in whom similar limitations of other calculation methods will equally apply.

REFERENCES

1. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol*. 2017;82:128-136.
2. Bellamy N, Hochberg M, Tubach F, et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. *Arthritis Care Res*. 2015;67(7):972-980.
3. Byrd JWT, Jones KS, Gwathmey W. Femoroacetabular impingement in adolescent athletes: outcomes of arthroscopic management. *Am J Sports Med*. 2016;44(8):2106-2111.
4. Carton PF, Filan D. Labral cuff refixation in the hip: rationale and operative technique for preserving the chondrolabral interface for labral repair. A case series. *J Hip Preserv Surg*. 2018;5(1):78-87.
5. Carton PF, Filan DJ. The clinical presentation, diagnosis and pathogenesis of symptomatic sports-related femoroacetabular impingement (SRFAI) in a consecutive series of 1021 athletic hips. *Hip Int*. 2019;29(6):665-673.
6. Chahal J, Thiel GSV, Mather RC, Lee S, Salata MJ, Nho SJ. The minimal clinical important difference (MCID) and patient acceptable symptomatic state (PASS) for the modified Harris Hip Score and Hip Outcome Score among patients undergoing surgical treatment for femoroacetabular impingement. *Orthop J Sports Med*. 2014; 2(2)(suppl):2325967114S00105.
7. 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 Rev*. 2018;6(9):e2.
8. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimally clinical important difference: a review of concepts and methods. *Spine J*. 2007;7(5):541-546.
9. Cvetanovich GL, Weber AE, Kuhns BD, et al. Hip arthroscopic surgery for femoroacetabular impingement with capsular management: factors associated with achieving clinically significant outcomes. *Am J Sports Med*. 2018;46(2):288-296.
10. Domb BG, Philippon MJ, Giordano BD. Arthroscopic capsulotomy, capsular repair and capsular plication of the hip: relation to atraumatic instability. *Arthroscopy*. 2013;29(1):162-173.
11. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149-158.
12. Gilmer BB, Comstock BA, Jette JL, Warme WJ, Jackine SE, Matsen FA. The prognosis for improvement in comfort and function after the ream-and-run arthroplasty for glenohumeral arthritis: an analysis of 176 consecutive cases. *J Bone Joint Surg Am*. 2012;94:e102.
13. Griffin DR, Dickenson EJ, O'Donnell J, et al. The Warwick Agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement. *Br J Sports Med*. 2016;50: 1169-1176.
14. Hart DL. Test-retest reliability of an abbreviated self-report overall health status measure. *J Orthop Sports Phys Ther*. 2003;33(12): 734-744.
15. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993;2(3):217-227.
16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.
17. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopaedic conditions. *J Orthop Surg Res*. 2015;10:24.
18. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient reported outcome measures for hip arthroscopic surgery. *Am J Sports Med*. 2013;41(9):2065-2073.
19. Khan M, Habib A, de Sa D, et al. Arthroscopy up to date: femoroacetabular impingement. *Arthroscopy*. 2016;32(1):177-189.
20. Krych AJ, Thompson M, Knutson Z, Scoon J, Coleman SH. Arthroscopic labral repair versus selective labral debridement in female patients with femoroacetabular impingement: a prospective randomized study. *Arthroscopy*. 2013;29(1):46-53.
21. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis*. 2007;66(suppl 3):iii40-iii41.
22. Larson CM, Giveans MR. Arthroscopic debridement versus refixation of the acetabular labrum associated with femoroacetabular impingement. *Arthroscopy*. 2009;25(4):369-376.

23. Levy DM, Kuhns BD, Chahal J, Philippon MJ, Kelly BT, Nho SJ. Hip arthroscopy outcomes with respect to patient acceptable symptomatic state and minimal clinically important difference. *Arthroscopy*. 2016;32(9):1877-1886.
24. Mann BJ, Gosens T, Lyman S. Quantifying clinically significant change: a brief review of methods and presentation of a hybrid approach. *Am J Sports Med*. 2012;40(10):2385-2393.
25. Nepple JJ, Philippon MJ, Campbell KJ, et al. The hip fluid seal, part II: the effect of an acetabular labral tear, repair, resection, and reconstruction on hip stability to distraction. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(4):730-736.
26. Nho SJ, Magennis EM, Singh CK, Kelly BT. Outcomes after arthroscopic treatment of femoroacetabular impingement in a mixed group of high-level athletes. *Am J Sports Med*. 2011;39(1)(suppl): 14-19.
27. 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(5):582-592.
28. Nwachukwu BU, Chang B, Adjei J, et al. Time required to achieve minimal clinically important difference and substantial clinical benefit after arthroscopic treatment of femoroacetabular impingement. *Am J Sports Med*. 2018;46(11):2601-2606.
29. Nwachukwu BU, Chang B, Fields K, et al. Defining the "substantial clinical benefit" after arthroscopic treatment of femoroacetbular impingement. *Am J Sports Med*. 2017;45(6):1297-1303.
30. Nwachukwu BU, Fields K, Chang B, Nawabi DH, Kelly BT, Ranawat AS. Preoperative outcome scores are predictive of achieving the minimal clinically important difference after arthroscopic treatment of femoroacetabular impingement. *Am J Sports Med*. 2017;45(3): 612-619.
31. Potter BK, Freedman BA, Anderson RC, Bojescul JA, Kulo TR, Murphy KP. Correlation of Short Form-36 and disability status with outcomes of arthroscopic acetabular labral debridement. *Am J Sports Med*. 2005;33(6):864-870.
32. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Res Ther*. 2015;17(1):143.
33. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch Intern Med*. 1993;153:1337-1342.
34. Ross JR, Larson CM, Adeoye O, Kelly BT, Bedi A. Erratum to: Residual deformity is the most common reason for revision hip arthroscopy: a three-dimensional CT study. *Clin Orthop Relat Res*. 2015;473(3): 1167.
35. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain*. 2004;8(4):283-291.
36. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the health utilities index mark II. *Pharmacoeconomics*. 1999;15(2):141-155.
37. Sansone M, Ahldén M, Jonasson P, et al. Good results after hip arthroscopy for femoroacetabular impingement in top-level athletes. *Orthop J Sports Med*. 2015;3(2):2325967115569691.
38. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63(5): 524-534.
39. Thorborg K, Kraemer O, Madsen AD, Holmich P. Patient-reported outcomes within the first year of hip arthroscopy and rehabilitation for femoroacetabular impingement and/or labral injury: the difference between getting better and getting back to normal. *Am J Sports Med*. 2018;46(11):2607-2614.
40. Tubach F, Ravaud P, Martin-Mola E, et al. Minimum clinically important improvement and patient acceptable symptomatic state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res*. 2012; 64(11):1699-1707.
41. Weber AE, Kuhns BD, Cvetanovich GL, Grzybowski JS, Salata MJ, Nho SJ. Amateur and recreational athletes return to sport at a high rate following hip arthroscopy for femoroacetabular impingement. *Arthroscopy*. 2017;33(4):748-755.
42. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*. 2012;20(3):160-166.