ORIGINAL REPORT

RESPONSIVENESS OF THE ACTIVITIES OF DAILY LIVING SCALE OF THE KNEE OUTCOME SURVEY AND NUMERIC PAIN RATING SCALE IN PATIENTS WITH PATELLOFEMORAL PAIN

Sara R. Piva, PhD, PT, OCS, FAAOMPT[1], Alexandra B. Gil, MS, PT[1], Charity G. Moore, PhD, MSPH[2] and G. Kelley Fitzgerald, PhD, PT, OCS[1]

From the [1]Department of Physical Therapy, School of Health and Rehabilitation Science, and [2]Department of Medicine, Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, USA

Objective: **To assess internal and external responsiveness of the Activity of Daily Living Scale of the Knee Outcome Survey and Numeric Pain Rating Scale on patients with patellofemoral pain.**
Design: **One group pre-post design.**
Subjects: **A total of 60 individuals with patellofemoral pain (33 women; mean age 29.9 (standard deviation 9.6) years).**
Methods: **The Activity of Daily Living Scale and the Numeric Pain Rating Scale were assessed before and after 8 weeks of physical therapy program. Patients completed a global rating of change scale at the end of therapy. The standardized effect size, Guyatt responsiveness index, and the minimum clinical important difference were calculated.**
Results: **Standardized effect size of the Activity of Daily Living Scale was 0.63, Guyatt responsiveness index was 1.4, area under the curve was 0.83 (95% confidence interval: 0.72, 0.94), and the minimum clinical important difference corresponded to an increase of 7.1 percentile points. Standardized effect size of the Numeric Pain Rating Scale was 0.72, Guyatt responsiveness index was 2.2, area under the curve was 0.80 (95% confidence interval: 0.70, 0.92), and the minimum clinical important difference corresponded to a decrease of 1.16 points.**
Conclusion: **Information from this study may be helpful to therapists when evaluating the effectiveness of rehabilitation intervention on physical function and pain, and to power future clinical trials on patients with patellofemoral pain.**

Key words: knee, minimum clinical important difference, area under the curve, standardized effect size.

J Rehabil Med 2009; 41: 129–135

INTRODUCTION

Patellofemoral pain (PFP) is a common musculoskeletal condition, accounting for 20–40% of all knee problems in adolescents and active young adults (1, 2). PFP is characterized by anterior knee pain and crepitation in the patellofemoral joint during and after weight-bearing activities such as walking up/down stairs, squatting and running. Pain while sitting with the knees flexed, occasional weakness, giving way and catching sensations are also characteristics of PFP (3).

Usual goals of rehabilitation interventions in patients with PFP are to reduce pain and improve physical function. To assess the achievement of these goals, clinicians use self-reported measures of pain and function throughout the process of care. Some of the commonly used measures of pain and function in these patients are the Numeric Pain Rating Scale (NRPS) and the Activity of Daily Living Scale (ADLS) of the Knee Outcome Survey, respectively. The NPRS is a reliable and valid measure of pain intensity (4–6), which in our clinical environment is administered at each rehabilitation visit. The ADLS is a reliable and valid self-reported knee specific measure of physical function (7, 8), which in our clinical setting is administered once a week to follow-up on patient's progression. Although these instruments are commonly used in patients with PFP (9–12), their responsiveness in this population have not been reported.

Responsiveness evaluates the ability of a measure accurately to detect change in patients' health status over time when change has occurred (13, 14). Husted et al. (15) suggested 2 major aspects of responsiveness: internal and external responsiveness. The usefulness of assessing internal or external responsiveness will depend on how clinicians intend to use the measure.

Internal responsiveness is characterized by the ability of a group measure to change over time in response to an intervention. It is usually measured within the context of randomized clinical trials or repeated measures designs and will depend on the intervention and measure used. Internal responsiveness statistics are based on the distribution of the data and is built upon the statistical properties of a study's results and include indices such as the standardized effect size and the Guyatt Responsiveness Index (16, 17).

External responsiveness reflects the association between individual changes in a measure over time and the corresponding individual changes in an external reference measure of health status. As it assesses changes at an individual level, quantification of change considers whether persons are deemed better or worse based on the external reference. For that reason, unlike internal responsiveness, it will not depend

on the intervention and measure used, but rather on the choice of external reference (15). External responsiveness method defines clinically meaningful change as whether the patient has actually improved in the external reference amount that is likely to be perceived as important to the patient (16). The receiver operating characteristic (ROC) curve is a method used to assess external responsiveness and can be used to determine the minimum clinical important difference (MCID). The MCID is defined as the smallest change required in a given outcome that is considered to be worthwhile or important to a patient (17, 18). Interpretation of external responsiveness statistics such as the MCID is attractive because, as it characterizes changes at the individual patient level, in another study of similar patients, the same relationship should be observed and the results of different studies can be compared (15). Therefore, knowing the MCID values for the NPRS and the ADLS would help clinicians to determine whether the magnitude of change in these measures in response to rehabilitation interventions could be considered clinically meaningful.

Although some studies investigated the internal responsiveness of the ADLS in patients with knee conditions (7, 8), to the best of our knowledge no study attempted to determine the external responsiveness of the ADLS in patients with PFP. One of the studies that investigated the internal responsiveness of the ADLS comprised only 20% of subjects with PFP and 57% of the population underwent surgical intervention (7). In the other study 36% of the population had PFP and 43% of the population underwent surgery (8). We found no studies that investigated the responsiveness of the NPRS in patients with PFP. Therefore, the purpose of this study is to assess the external as well as the internal responsiveness of the ADLS and NPRS in patients with PFP. We seek to provide benchmarks against which to compare outcomes in future intervention studies with this population as well as to provide effect sizes for power analyses.

METHODS

Patients

Subjects were participants in a multicenter study that investigated the association of physical impairments and functional outcome in subjects with PFP syndrome who underwent physical therapy treatment. Four physical therapy clinics located throughout the USA participated in this study (Minot Air Force Base, Minot, ND; Lackland Air Force Base, San Antonio, TX; Travis Air Force Base, Fairfield, CA; and University of Pittsburgh's Centers for Rehab Services, Pittsburgh, PA, USA). The study was approved by each site's Institutional Review Board and all subjects provided consent before participation.

Individuals between the ages of 12 and 50 years referred to physical therapy for treatment of PFP were invited to participate in this study if they met the following inclusion criteria: primary complaint of PFP, pain in one or both knees that was aggravated with physical activities, duration of signs and symptoms greater than 4 weeks, history of insidious onset of pain not related to trauma, and pain in the patellar region with *at least 3 out of the following:* manual compression of the patella against the femur at rest or during an isometric knee extensor contraction, palpation of the postero-medial and postero-lateral borders of the patella, resisted isometric quadriceps femoris muscle contraction, squatting, stair climbing, kneeling, or prolonged sitting.

Exclusion criteria included previous patellar dislocation, knee surgery over the past 2 years, concomitant diagnosis of peripatellar

bursitis or tendonitis, internal knee derangement, systemic arthritis, ligamentous knee injury or laxity, plica syndrome, Sinding Larsen's disease, Osgood Schlatter's disease, infection, malignancy, musculoskeletal or neurological lower extremity involvement that interferes with physical activity, and pregnancy.

This study reports on 60 patients (33 women, mean age 29.9 (standard deviation 9.6) years) who completed the 2-month follow-up testing (Table I). These 60 patients represent 81% of the total original enrollment of 74 patients. Baseline characteristics of the patients who completed and the ones who dropped out of the study before the 2-month follow-up were not statistically different nor appeared to be clinically meaningful for the variables gender, age, height, weight, race, type of work, use of pain medication, chronicity of pain, level of physical activity, and ADLS and NPRS scores (Pearson χ^2 test used for nominal variables, Mann-Whitney U- or independent sample t-tests used for continuous variables depending on data distribution). Reasons for drop-outs included: 5 patients had job-related time constraints, 3 patients could not be contacted for the 2-month assessment, 2 patients suffered major knee trauma due to sports, 2 patients had no more pain after baseline assessment, one patient had a spinal injury, and one had a family health issue.

Procedures

Baseline examination included the assessment of pain intensity and physical function using the NPRS and ADLS, respectively. All subjects then underwent the same physical therapy exercise program. The exercise program incorporated 8 weeks of strengthening exercises, stretching exercises, and patellar taping, and was based on prior evidence in the literature suggesting that each of these treatment elements improve pain and function in patients with PFPS (19–23). Because this is not an intervention paper, the intervention is summarized briefly: patellar taping was applied at the beginning of each treatment session as originally proposed by McConnell (21). Next, a warm-up took place by having the patient ride a stationary bicycle for 5 min. Following the warm-up, the stretching exercises included quadriceps, hamstrings, and plantar flexors stretching. Strengthening exercises included quadriceps strengthening in weight-bearing (double leg squats, and unilateral step-down and step-up) and non-weight-bearing conditions (quadriceps settings, straight leg raises, and short arc leg extension). Outcome measures were repeated at 2-month follow-up after completion of the physical therapy program.

Table I. *Baseline characteristics of study population (n = 60)*

Variables	Means (SD), or frequency (%)
Age, years	29.9 (9.6)
Females, *n* (%)	33 (55)
Height, cm	170 (10)
Weight, kg	74.8 (15.6)
Ethnicity, *n* (%)	
Caucasian	41 (69)
African-American	7 (11)
Hispanic	7 (11)
Asian	2 (4)
Other	3 (5)
Chronicity of pain, months (%)	
1–3	23 (38)
4–6	13 (22)
7–12	6 (10)
13–24	11 (18)
>25	7 (12)
Baseline ADLS score (possible scores 0–100)	67 (15.5)
Baseline Worst NPRS score (possible scores 0–10)	5.5 (2.3)

SD: standard deviation; ADLS: activities of daily living scale; NPRS: numeric pain rating scale.

Measures

The following self-report measures were used for the purposes of this analysis:

- ADLS: this knee-specific measure of physical function consists of 14 items. Six items assess knee symptoms and 8 items assess functional limitation during the performance of daily activities (7). Each item is scored on 6-point Likert scale (0–5 points). The ADLS score is transformed to a 0–100 point scale with 100 indicating the absence of symptoms and functional limitations. Patients completed the ADLS at baseline and 2-month follow-up.

- NPRS: an 11-point numeric scale was anchored on the left with the phrase "No Pain" and on the right with the phrase "Worst Imaginable Pain" (5). Subjects rated their worst level of pain in the past 24 h at baseline and at the 2-month follow-up. We decided to calculate responsiveness of the NPRS using the worst pain rather than the current or least amount of pain during the last 24 h because a considerable proportion of patients with PFP report very low current and least pain levels. Therefore, the utilization of either the current or least amount of pain to determine responsiveness may be problematic due to the potential floor effects of these measures. Furthermore, in a busy clinical environment having a single measure of pain to assess improvement is more convenient than having several measures or an average of them.

- Global rating of change: at the 2-month follow-up, patients were asked to rate their overall change in clinical status since the beginning of physical therapy treatment using a 15-point rating scale (18). The global rating of change has been used in research as an outcome measure as well as an external anchor to compare outcome measures (24–26). The global rating of change ranges from +7 ("a very great deal better") to 0 ("about the same") to –7 ("a very great deal worse"). Intermittent descriptors of improving are assigned values from +1 to +6, and of worsening are assigned values from –1 to –6. Patients with a rating of +3 ("somewhat better") or higher were considered to have improved. Patients with an average rating of +2 ("a little bit better") to –2 ("a little bit worse") were considered to have minimally or not changed. Patients with a rating of –3 ("somewhat worse") or lower were considered to have worsened. We have conservatively chosen to place patients who described to be a little bit better or worse (+2 or –2, respectively) into the minimal or no change group because we wanted to be confident that the patients in the improved or worsened groups would have most likely experienced a considerable change.

Statistics

Analyses were performed with SPSS version 14.0 (SPSS Science Inc., Chicago, IL, USA). The distribution of changes in ADLS and NPRS of the overall group, as well as the improved, minimal/no change, and worsened subgroups, were tested for normality using the Kolmogorov-Smirnov test (2-tailed). Internal responsiveness of the ADLS and the NPRS was first characterized by calculating Standardized Effect Size and the Guyatt Responsiveness Index at the 2-month follow-up (14, 27). Standardized effect size was calculated as the change score on the measure (difference between the mean baseline scores and follow-up scores for each individual) divided by the standard deviation of baseline scores (13, 28). Because only 5 subjects worsened, we present only the descriptive characteristics rather than effect sizes of this subgroup. The Guyatt Responsiveness Index was calculated as the ratio of mean change of patients who reported improvement divided by the standard deviation of the change of patients reporting minimal or no change based on the global rating of change (14).

External responsiveness of the ADLS and NPRS was characterized by calculating the area under the ROC curve and its 95% confidence interval (95% CI) (15, 29). The area under the curve can be used as a quantitative method for assessing a scale's ability to distinguish patients who have improved from those who have minimally or not changed based on the global rating of change (improved patients had a global rating of 3 or above while minimally or not improved ones had global rating of 2 or below). For the ROC curve calculation the patients who have minimally or not changed were collapsed with the patients who worsened. Therefore, the areas under the ROC curves assess the ability of the ADLS and NPRS to distinguish patients who improved from those who have not, while the small sample of patients who worsened did not allow construction of ROC curve to distinguish patients who have worsened from the ones who have not. As a general rule, areas under the curve between 0.7 and 0.8 are considered to have acceptable discrimination; areas under the curve from 0.8 to 0.9 are considered to have excellent discrimination; and areas under the curve above 0.9 are considered to have outstanding discrimination (30). The point of the ROC curve on the upper-most left-hand corner was used to estimate the MCID. The MCID represents the point with the highest sensitivity (probability of the measure correctly classifying patients who demonstrate change on the global rating of change) and specificity (probability of the measure correctly classifying patients who have minimally or not changed on the global rating of change) in the ROC curve (31, 32).

Table II. *Descriptive statistics and responsiveness characteristics for the Activities of Daily Living Scale (ADLS) and Numeric Pain Rating Scale (NPRS)*

	ADLS*			NPRS†		
	Baseline	2- month	Change	Baseline	2-month	Change
Mean change (SD)						
Overall group (*n*=60)	67.0 (15.5)	76.7 (18.8)	9.7 (15.8)	5.5 (2.3)	3.8 (2.7)	–1.7 (2.8)
Improved group (*n*=36)	69.5 (14.2)	86.4 (12.5)	16.9 (13.2)	5.1 (2.3)	2.3 (2.0)	–2.8 (2.8)
Minimal or no change group (*n*=19)	62.6 (17.6)	63.1 (18.7)	0.5 (12.1)	6.1 (2.3)	5.5 (2.1)	–0.6 (1.5)
Worsened group (*n*=5)	65.5 (15.2)	58.3 (11.5)	–7.2 (16.9)	6.0 (2.1)	7.8 (0.8)	1.8 (1.5)
Standardized effect size‡						
Overall group (*n*=60)		0.63			0.74	
Improved group (*n*=36)		1.19			1.22	
Minimal or no change group (*n*=19)		0.03			0.26	
Guyatt Responsiveness Index§ (*n*=55)		1.4			1.9	
Minimum clinically important difference¶ (*n*=60)		7.14			–1.16	
Area under the ROC curve (95% CI)		0.83 (0.73, 0.94)			0.84 (0.70, 0.92)	

*Negative values represent decreased physical function.
†Negative values represent decreased pain.
‡Standardized Effect Size = (postoverall – preoverall)/SD preoverall.
§Guyatt Responsiveness Index = (postimprovers – preimprovers)/SD post – postminimal or no change – preminimal or no change.
¶Minimum clinically important difference was derived from the ROC curve.
SD: standard deviation, ROC: receiver operating characteristic; CI: confidence interval.

RESULTS

At the 2-month follow-up examination 36 subjects (60%) were classified as having improved, 19 (32%) as having minimally or not changed, and 5 as having worsened (8%) based on the global rating of change (Table II). The distribution of data for changes in ADLS and NPRS and the global rating of change are depicted in Fig. 1. The distribution of data for changes in ADLS and NPRS of the overall and subgroups that improved, minimally/not changed, and worsened did not depart from normality (*p*-value of Kolmogorov-Smirnov Z > 0.4). Table II illustrates the descriptive statistics of changes in ADLS and NPRS, standardized effect sizes, Guyatt responsiveness index, area under the ROC curve with its 95% CI, and MCID for the ADLS and the NRPS.

Assessment of the responsiveness of the ADLS demonstrated a moderate standardized effect size of 0.63 (overall), and a Guyatt responsiveness index of 1.4. The area under the curve was 0.83 (95% CI: 0.72, 0.94), which is considered as having excellent discrimination (30). The MCID for functional improvement corresponded to a change of 7.1 percentage points,

which represents 5 points change in the raw ADLS score. Responsiveness of the NPRS demonstrated a moderate overall standardized effect size of 0.74 and a Guyatt responsiveness index of 1.9. The area under the curve was 0.84 (95% CI: 0.70, 0.92), also considered excellent discrimination (30). The MCID for decreases in pain corresponded to a change of −1.2 points. The ROC curves are depicted in Fig. 2.

DISCUSSION

Our results for external responsiveness indicate that an increase of at least 7.1 percentage points in the ADLS and a decrease of at least 1.2 points in the NPRS in patients with PFP represent clinically meaningful improvements in physical function and patient's perceived level of pain, respectively. The MCID values of the ADLS and NPRS represent the change score in these measures that best classifies patients as improved (15). When interpreting change in patient's status based on these values, it is important to acknowledge that these values represent the minimum change, rather than moderate or larger
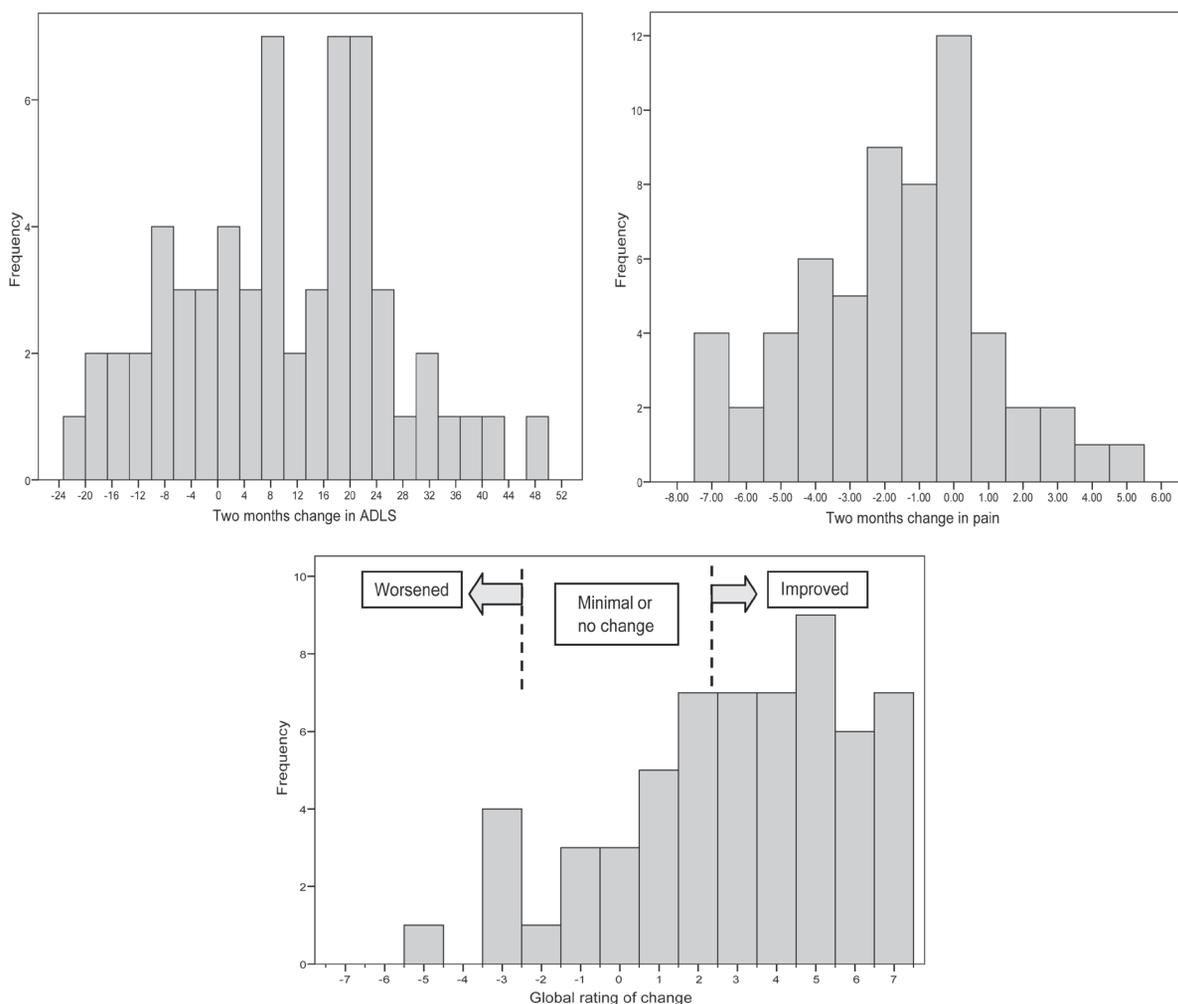

Fig. 1. Distribution of frequencies in the Activities of Daily Living Scale (ADLS), Numeric Pain Rating Scale (NPRS), and Global Rating of Change.
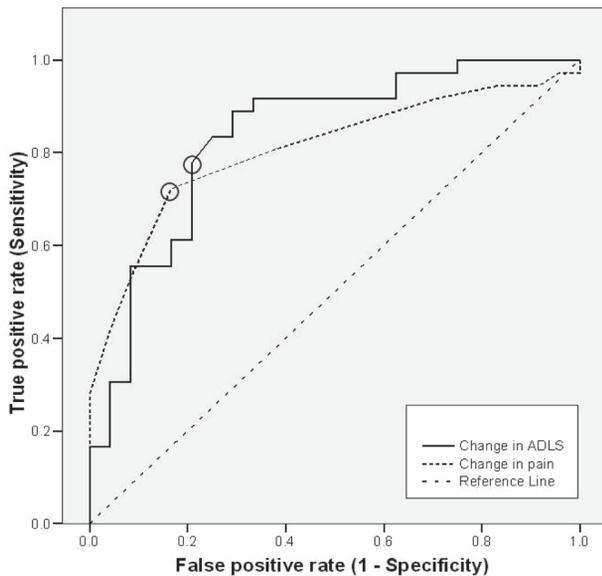
Fig. 2. Receiver operating characteristic curves for the changes in Activities of Daily Living Scale (ADLS) and Numeric Pain Rating Scale (NPRS). The circled values are the point nearest the uppermost left-hand corner of the graph for each measure. These points represent the minimum clinical important difference for the ADLS (7.14 percentage points) and NPRS (1.16 points).

changes, required in the ADLS and NPRS that is considered to be worthwhile or important to patients with PFP. For clinical purposes integer values are more convenient for clinicians and patients. Therefore, due to the small magnitude of the decimal values of the MCID, we suggest rounding the MCID values to the next lowest values (7 percentage points for the ADLS and 1 point for the NPRS). The interpretation of these MCID values are attractive because in another study of similar patients, the same relationship should be observed and the results of different studies can be compared (15).

The areas under the ROC curves were 0.83 and 0.84 for the ADLS and NPRS, respectively. For the ADLS, this means that if we select 2 patients at random, one with improvement in function and the other without, the probability is 0.83 that the patient with improvement will have a higher change on the ADLS than the patient without improvement (33). The same interpretation would apply to the NPRS. Because the lower bounds of the 95% CI of the area under the curve for these measures were above 0.70, we believe that even in the worst case scenario the discriminatory accuracy of both instruments are adequate. Area under the curve of at least 70% has been interpreted as having acceptable discriminatory accuracy (30).

Because the MCID of the NPRS has not been reported in patients with PFP, we compared our MCID value of 1.2 points with the ones reported in clinical trials of patients with spine-related problems or chronic pain conditions (26, 32, 34). Childs et al. (26) studied responsiveness of the NPRS in a population of patients with primary low back pain who received physical therapy. They reported MCIDs for decreases in pain of 2.2 and 1.5 for the 1-week and 4-week follow-up,

respectively. Grotle et al. (32) determined the responsiveness of the NPRS in 2 groups of patients at different time-points: at 4-week follow-up for patients with acute low back pain, and at 3-month follow-up for patients with chronic low back pain. They reported MCIDs of 1.5 points and 0.5 points for the groups with acute and chronic low back pain respectively. Farrar et al. (34) examined data for patients enrolled in 10 trials of chronic pain that used the same study design and procedures. Chronic pain condition of their sample of trials included diabetic neuropathy (3 trials), post-herpetic neuralgia (3 trials), chronic low back pain (2 trials), fibromyalgia (1 trial), and hip or knee osteoarthritis (1 trial). In their report the average length of studies follow-up was 8 weeks. They reported an MCID of 1.7 points (34). Therefore, independently of the musculoskeletal condition, it seems that the MCID value in our study is similar to the ones reported in studies that used a similar follow-up time (2 months).

To the extent of our knowledge this is the first time the MCID has been reported for the ADLS. External responsiveness was not assessed in the 2 previous studies on the responsiveness of the ADLS (7, 8). However, external responsiveness is not the only way to assess clinical significant differences. Methods to evaluate clinical significant differences can be either anchor-based or distribution-based (16). Anchor-based methods use an external reference standard such as the global rating of change to calculate cut-offs in scores that represent meaningful changes for the patients. An example is the calculation of the MCID. Distribution-based methods rely on statistical properties such as variances and error in the measurement to calculate cut-offs in scores that represent certain level of statistical confidence that the change exceeds the bounds of measurement error (29, 35). An example is the calculation of the minimum statistical change. Therefore, the minimum statistical change is used to evaluate statistically meaningful levels of change rather than clinically meaningful levels of change like the MCID.

Because the 2 previous studies on the responsiveness of the ADLS (7, 8) did not use a external reference to calculate MCID, and our study was not designed to determine error in the measurement (reliability) to determine the minimum statistical change of the ADLS, we decided to compare our MCID value (anchor-based method) with the meaningful statistical change of these studies (distribution-based method). We calculated the minimum statistical change of the ADLS for each study as $1.96 \times$ standard error of the measurement (35). The standard error of the measurement of Irrgang et al.'s study (7) was 3.2 points, whereas the standard error of the measurement of Marx et al.'s study (8) was 4.8 points (values calculated from their reported intraclass correlation coefficient and standard deviation). The minimum statistical change for the studies by Irrgang et al. (7) and Marx et al. (8) were 6.2 and 9.5, respectively. Therefore, the clinically meaningful level of change in our study of 7.14 points was similar to the statistically meaningful levels of change of 6.2 and 9.5 of their studies. Even though the MCID and the minimum detectable change are calculated upon different statistical constructs, the similarity of these values improves the confidence on the MCID value of the ADLS (6).

Prior authors have suggested that combination of change scores values calculated from distribution- and anchor-based methods should be taken collectively to support the identification of change scores that are clinically meaningful (36).

The results of internal responsiveness (standardized effect size and Guyatt responsiveness index) and the descriptive statistics reported in Table II are helpful in 3 ways: (*i*) to establish expected levels of change after rehabilitation in patients with PFP who have similar characteristics as those enrolled in this study; (*ii*) to calculate sample size in future clinical trials of rehabilitation in patients with PFP when the ADLS and NPRS are the primary outcomes; and (*iii*) to qualitatively compare which measure (ADLS or NRPS) is more "internally responsive". Our study demonstrated that the internal responsiveness of the ADLS and NRPS are very similar. For both instruments the standardized effect size values for the overall and improved groups are considered moderate and large, respectively (37). Although there is no consensus in the literature for the interpretation of the magnitude of Guyatt responsiveness index values, the Guyatt responsiveness index for both measures were very similar. Husted et al. (15) stated that the use of internal responsiveness statistics for comparisons across studies is difficult because the calculations are specific to each study and there is no well-defined interpretation that can be given to particular values of the statistics, independent of study design. Furthermore, as internal responsiveness also depends on the intervention used to determine treatment efficacy, the effect sizes reported in this study should not be used with other populations such as patients with PFP who underwent surgery, or patients with PFP who received a different rehabilitation intervention.

Although some may question the appropriateness of applying parametric treatment to total scores derived from multi-item scales such as the ADLS, we believe the parametric approach we used for internal responsiveness was appropriate. The main assumptions about using parametric methods such as means, standard deviations, and effect sizes are that: (*i*) the data approximately follow some distribution (usually a normal distribution), and (*ii*) the level of measure has interval-level property (38). With regards to the first assumption, we did test the distribution of the change data for all the subgroups of patients and the data did not depart from the normal distribution. With regards to the second assumption, the use of total scores of multi-item scales (in which each individual item is ordinal data) as an interval-level measure has been supported given that there is a large range of scores with sufficient distinct values (39). The ADLS derives from 14 items, each scored on 6-point Likert scale (0–5 points) and the score is transformed in a range of 0–100. For our analysis, change in the ADLS generated 49 distinct values.

We believe the results of this study can be applied to the general population of patients with PFP who are selected to receive rehabilitation. Although we used a convenience sample, the characteristics of our sample are similar to the ones reported in other studies in patients with PFP who underwent conservative interventions (19–23). The information reported here can potentially be used as benchmarks against which to compare outcomes in future intervention studies with this particular population as well as to provide effect sizes for power analyses of similar rehabilitation interventions. Because we recruited subjects who were referred to conservative intervention, and also because prior surgery (within 2 years) was an exclusion criterion, it is possible that these patients represent a group of individuals who respond differently to rehabilitation from those referred to surgery. As such, although the external responsiveness is dependent on the choice of the external reference rather than on the intervention used (15), we recommend caution in using the MCID in surgical patients. The minimum change required in the ADLS and NPRS that is considered to be worthwhile or important to surgical patients may not be the same. A limitation of this study may be that we only asked patients to complete the global rating of change, rather than asking patients and clinicians. In content areas such as functional gain and pain, patient's rating regarding their perception of change over short period of time appears to be a good selection of external anchor, whereas clinician's ratings may not be consistent (36). In addition, because the condition of a very small number of patients worsened, we could not calculate the MCID for worsening, which should not be assumed to have the same value as the MCID for improvement.

In conclusion, in patients with PFP, a 7 percentage-point increase on the ADLS and 1-point decrease on the NPRS seems to represent the minimum clinically meaningful improvements in these measures. Therapists could use the information from this preliminary report to evaluate the effectiveness of rehabilitation intervention on physical function and pain and to power future clinical trials on patients with PFP.

ACKNOWLEDGMENTS

REFERENCES

1. Insall J. Current concepts review: patella pain. J Bone Joint Surg 1982; 64A: 147–152.
2. Brody LT, Thein JM. Nonoperative treatment for patellofemoral pain. J Orthop Sports Phys Ther 1998; 28: 336–344.
3. Thomee R, Augustsson J, Karlsson J. Patellofemoral pain syndrome: a review of current issues. Sports Med 1999; 28: 245–262.
4. Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999; 79: 231–252.
5. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994; 58: 387–392.
6. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. Spine 2007; 32: 580–585.
7. Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. J Bone Joint Surg Am 1998; 80: 1132–1145.

8. Marx RG, Jones EC, Allen AA, Altchek DW, O'Brien SJ, Rodeo SA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. J Bone Joint Surg Am 2001; 83-A: 1459–1469.

9. Gerbino PG, Griffin ED, d'Hemecourt PA, Kim T, Kocher MS, Zurakowski D, et al. Patellofemoral pain syndrome: evaluation of location and intensity of pain. Clin J Pain 2006; 22: 154–159.

10. Wilson T, Carter N, Thomas G. A multicenter, single-masked study of medial, neutral, and lateral patellar taping in individuals with patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33: 437–443; discussion 444–448.

11. Piva SR, Goodnite E, Childs JD. Strength around the hip and flexibility of soft tissues in individuals with and without patellofemoral pain syndrome. J Orthop Sports Phys Ther 2005; 35: 793–801.

12. Karataglis D, Green MA, Learmonth DJ. Functional outcome following modified Elmslie-Trillat procedure. Knee 2006; 13: 464–468.

13. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000; 25: 3192–3199.

14. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40: 171–178.

15. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000; 53: 459–468.

16. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res 1993; 2: 221–226.

17. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000; 6: 39–49.

18. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989; 10: 407–415.

19. Salsich GB, Brechter JH, Farwell D, Powers CM. The effects of patellar taping on knee kinetics, kinematics, and vastus lateralis muscle activity during stair ambulation in individuals with patellofemoral pain. J Orthop Sports Phys Ther 2002; 32: 3–10.

20. Witvrouw E, Lysens R, Bellemans J, Peers K, Vanderstraeten G. Open versus closed kinetic chain exercises for patellofemoral pain. A prospective, randomized study. Am J Sports Med 2000; 28: 687–694.

21. McConnell J. The management of chondromalaciae patellae: a long-term solution. Aust J Physiother 1986; 32: 215–233.

22. McMullen W, Roncarati A, Koval P. Static and isokinetic treatments of chondromalacia patella: A comparative investigation. J Orthop Sports Phys Ther 1990; 12: 256–266.

23. Stiene HA, Brosky T, Reinking MF, Nyland J, Mason MB. A comparison of closed kinetic chain and isokinetic joint isolation exercise in patients with patellofemoral dysfunction. J Orthop Sports Phys Ther 1996; 24: 136–141.

24. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993; 20: 561–565.

25. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol 1994; 47: 81–87.

26. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine 2005; 30: 1331–1334.

27. Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. Spine 2003; 28: 607–615.

28. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27 Suppl 3: S178–S189.

29. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986; 39: 897–906.

30. Hosmer DW, Lemeshow S, editors. Applied logistic regression. 2nd edn. New York: John Wiley & Sons, Inc; 2000.

31. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996; 65: 71–76.

32. Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine 2004; 29: E492–E501.

33. Obuchowski NA. Receiving operating characteristic curves and their use in radiology. Radiology 2003; 229: 3–8.

34. Farrar JT, Young JP Jr, 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: 149–158.

35. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999; 37: 469–478.

36. Haley S, Fragala-Pinkham M. Interpreting change scores of tests and measures used in physical therapy. Phys Ther 2006; 86: 735–743.

37. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000; 6: 39–49.

38. Altman DG, editor. Practical statistics for medical research. London: Chapman and Hall; 1991.

39. Pett MA, editor. Nonparametric statistics for health care research. London: Sage Publications; 1997.