

# High relative reliability and responsiveness of the forgotten joint score-12 in patients with femoroacetabular impingement undergoing hip arthroscopic treatment. A prospective survey-based study

Ida Birgitte Bramming<sup>1\*</sup>, Signe Kierkegaard<sup>2</sup>, Bent Lund<sup>2</sup>,  
Stig S. Jakobsen<sup>3</sup> and Inger Mechlenburg<sup>3,4</sup>

<sup>1</sup>Department of Sports Science and Clinical Biomechanics, Faculty of Health Sciences, University of Southern Denmark, Denmark,

<sup>2</sup>Department of Orthopaedic Surgery, Horsens Regional Hospital, Denmark,

<sup>3</sup>Department of Orthopaedic Surgery, Aarhus University Hospital, Denmark and

<sup>4</sup>Department of Clinical Medicine, Centre of Research in Rehabilitation (CORIR), Aarhus University Hospital and Aarhus University, Denmark.

\*Correspondence to: I. B. Bramming. E-mail: bramming86@hotmail.com

Submitted 17 November 2018; Revised 28 February 2019; revised version accepted 5 May 2019

## ABSTRACT

The forgotten joint score-12 (FJS-12) may be an advantageous questionnaire in young patients with high hip function and a low level of pain. We investigated the reliability and the responsiveness of the FJS-12 in patients with femoroacetabular impingement undergoing hip arthroscopic treatment. Fifty patients were included in the reliability study and 34 patients were included in the responsiveness study. Test–retest reliability was assessed with intraclass correlation coefficient (ICC), standard error of measurement (SEM) and minimal detectable change (MDC). Responsiveness was assessed from testing correlations between the FJS-12 and the Copenhagen Hip and Groin Outcome Score (HAGOS) of the change score, effect size (ES) and standardized response mean (SRM). Floor and ceiling effect were defined as present if the number of patients obtaining the maximum (100) and minimum score (0) exceeded 15%. The relative reliability was high (ICC = 0.9, 95% CI: 0.8–0.9) and the absolute reliability was low (SEM = 11, MDC<sub>individual</sub> = 32, MDC<sub>group</sub> = 4.5). The responsiveness was high, and the change score was highly correlated with the subscale ‘pain’ from the HAGOS and moderately correlated with the subscale ‘ADL’. Furthermore, the FJS-12 exceeded or equalled the HAGOS subscales in ES and SRM. Below 15% of the patients scored the maximum or minimum score. The FJS-12 has high reliability, high responsiveness to change and shows no floor or ceiling effect.

## INTRODUCTION

Evaluation of outcomes in patients undergoing conservative or operative treatment of the hip has changed over the past decade. Traditionally, rating systems have focused on the clinician assessment of joint range of motion, joint stability or function [1]. Due to increased patient participation and expectations to the surgical or non-surgical treatment, patient-reported outcome (PRO) measurements have become an important instrument when the concept

being measured is best known by the patient, or best measured from the patient’s perspective [1, 2].

Recently, Behrend *et al.* developed a new joint-specific questionnaire, the forgotten joint score-12 (FJS-12) [3]. This new PRO differentiates from other PROs with emphasis on the patient’s awareness and ability to forget about their joint in everyday life [3]. In general, one is not aware of a healthy joint during the usual activities of daily living, and it can therefore be regarded as ‘forgotten’ [4].

From a patient's perspective the question of awareness is covering a broader range of pain symptoms, from a tiny unpleasant sensation to sudden pain. No matter if the pain is experienced as a nagging, a sharp stabbing or a throbbing sensation, it will produce awareness. This suggests that joint awareness may be a more discerning measure of patient outcomes than domains traditionally captured in PROs [5]. A more discerning measure as in a more discriminating or sensitive outcome measure because it is based on a different domain 'hip joint awareness' and because it has very little ceiling effect.

The FJS-12 has been validated in patients with hip or knee joint replacements and the findings are promising [3–11], but to the best of our knowledge, no assessment has yet been conducted in patients with femoroacetabular impingement syndrome (FAIS) [12]. Strong ceiling effects impair a PROs ability to differentiate between patients with good and excellent outcomes. Studies suggest that the FJS-12 differentiates effectively in highly functioning patient's groups, which matches patients with FAIS who are typically young and physically active [3, 13].

In this study, we intended to investigate the post-operative test–retest reliability and the pre- to post-operative level of responsiveness of the FJS-12 by comparing change scores to the Copenhagen Hip and Groin Outcome Score (HAGOS) in a population of patients with FAIS assigned for hip arthroscopic treatment [14]. Furthermore, we aimed to estimate floor and ceiling effect for the FJS-12. We hypothesized that the relative reliability and the responsiveness would be high and that the FJS-12 had no floor or ceiling effect [15]. As the FJS-12 has a different approach than other PROs it should not outperform already available ones but add to patient-reported information.

## MATERIALS AND METHODS

### Sample size

As recommended by de Vet *et al.* [16], we decided *a priori* to include 50 patients for the test–retest reliability study to achieve an intraclass correlation coefficient (ICC) of minimum 0.8 with 95% confidence interval (CI) at maximum 0.1. For the study of the responsiveness the same study sample was used. In the investigation we calculated the change score, the effect size (ES) and the standardized response mean (SRM), which are not dependent on sample size [17].

### Design

This study is a prospective study investigating the reliability and responsiveness of the FJS-12 in a cohort of patients with FAIS performed according to the COSMIN

guidelines [18–20]. In addition, the floor and ceiling effect was investigated in accordance with the quality criteria for measurement properties by Terwee *et al.* [15].

### Patients

Patients were recruited consecutively from December 2014 to February 2017 from the HAFAL-cohort (patients in Horsens and Aarhus with Femoral Acetabular Impingement) consisting of patients scheduled for hip arthroscopic treatment for FAIS at Horsens Regional Hospital, Denmark [19]. This cohort is followed prospectively to evaluate the outcome of arthroscopic treatment for FAIS. Patients diagnosed with cam and/or pincer impingement aged 18–53 years and assigned for hip arthroscopic treatment by one experienced surgeon (BL), at Horsens Regional Hospital, Denmark, was eligible for participation. Patients were operated in supine position through standard antero-lateral and mid-anterior portals. After a small interportal capsulotomy was created, a diagnostic round was accomplished from both portals and the relevant pathology was addressed. Labral tears were refixated with suture anchors. The number of anchors used for the repair depended on the quality of the labrum, and the size of the tear. In patients with a grade 4 acetabular chondral defect according, microfracture was performed unless the area was larger than 2–3 cm<sup>2</sup>. Bony deformities, such as pincer (deformity on the acetabular rim) and/or cam deformity (deformity on the femoral neck), were addressed by osteoplasty using a motorized burr. The end-result was controlled by assessing range of movement under direct vision and under image intensifier control [21].

Following surgery, the patients followed a standardized, homebased rehabilitation programme instructed by experienced physiotherapists [19].

### Measurements

A Danish version (Appendix 1) of the FJS-12 translated according to ISPOR (International Society for Pharmacoeconomics and Outcomes Research) guidelines was provided by the developer of the questionnaire, Dr Giesinger [22]. The translation process revealed only minor adjustments in wording according to the description of 'have your attention drawn to the joint' and 'choose the most appropriate answer for each question'. The back translated English version was approved by the developer of the questionnaire and by two of the authors (SSJ and IM). A debriefing interview was performed in 10 patients for linguistic validation of the Danish version of the FJS-12.

The FJS-12 consists of 12 questions scored from 1 to 5 (never to mostly) on a Likert scale, a high score indicating a good outcome i.e. a high degree of being able to forget

about the affected hip joint in daily life. Thus, the raw score ranges from 12 to 60, a high score indicating a good outcome i.e. a high degree of being able to forget about the affected joint in daily life.

In the statistical analysis the raw score is linearly transformed to a 0–100 scale with the equation: final score =  $100 - ((\text{sum}(\text{item01 to item12}) - 12)/48 * 100)$ .

Item selection was performed based on literature research and expert opinion choosing 20 items relevant to the average population undergoing joint arthroplasty. A pilot sample with 46 patients undergoing hip joint replacement or knee joint replacement selected 12 items from the 20 that was included in the final version of the FJS-12 [3].

The HAGOS was developed specifically for young and middle-aged physically active people [14]. In this study, the subscales pain and ADL were chosen as a reference to estimate the level of responsiveness, as they contain questions similar to the FJS.

#### Data collection

The study consisted of two parts; a reliability study and a responsiveness study. To examine the reliability of a PRO, the test–retest can be calculated by the extent of agreement and reproducibility between two repeated measurements. To test the reliability, patients were asked to complete the FJS-12 6 months after hip arthroscopic surgery and again 3 weeks later. The interval was close enough to avoid genuine changes in the hip symptoms and far enough apart to avoid memory effects [17]. To ensure a high degree of stability of the hip symptoms, patients were asked to mark their current hip function level on a visual analogue scale (VAS) at both time points. *A priori* it was decided, that patients with a change of >20/100 points were excluded as this was regarded as a clinical relevant change in patient symptoms [23]. Sixty-three patients were sent an email with a link to the Danish version of the FJS-12 (Fig. 1).

Responsiveness is the ability of a measure to detect change in a patient's condition over time. To investigate the level of responsiveness of the Danish version of the

FJS-12, 34 patients were invited and asked to complete the FJS-12 preoperatively and again 6 months postoperatively. These 34 patients were part of the study of the test–retest reliability as well.

#### Ethics

Patients in the study were all enrolled in the HAFAI-cohort study to which they gave informed consent in accordance with the Declaration of Helsinki before inclusion [19, 24]. The study was approved by the Central Denmark Region Committee on Biomedical Research Ethics (Journal No 1-10-72-239-14) and by the Danish Data Protection Agency (1-16-02-499-14).

#### Statistics

STATA 14.1 was used for all statistical analyses and a *P* values of <0.05 was considered significantly. The descriptive results were given as mean with standard deviation (SD) and range or as percentage as appropriate.

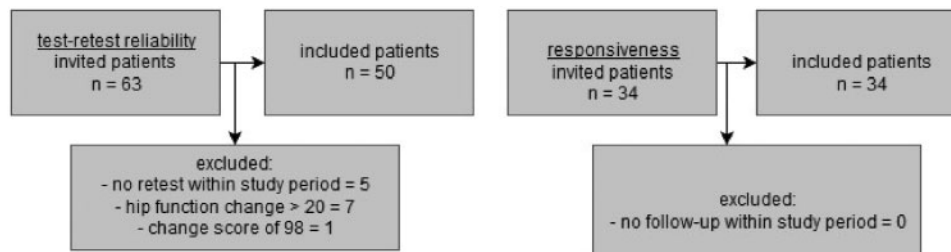
#### Missing data

Missing data were handled according to the scoring manual for the FJS-12 [3]. In case of missing answers, the missing values were replaced with the mean of the completed items. If more than four items were missing, the total score was treated as missing.

#### Test–retest reliability

Test–retest reliability is the extent to which an instrument is capable of measuring a variable with consistency [17]. In this study, the assessment was used to investigate whether patients score similarly on the FJS at two time points given that the patients' health condition remain unchanged.

The assumption of normal distribution was fulfilled, and Students' paired *t*-test was used to evaluate the differences between the two post-operative scores. The ICC with 95% CI was calculated using the two-way random effect model [21].



**Fig. 1.** Flow diagram presenting patients invited to participate and the number of patients excluded and included in the study of responsiveness and test–retest reliability, respectively.

The minimal detectable change (MDC) was used to define the amount of change in the variable that must be achieved to reflect a true difference. The MDC is the minimum amount of change in a patient's score that ensures the change is not the result of measurement error. MDC was calculated based on the standard error of measurement (SEM), which reflects the reliability of the response. The greater reliability of the measurement, the smaller the MDC [17].

The minimal detectable difference (MDC) at the individual level was calculated based on the SEM from the equation  $1.96 \cdot \sqrt{2} \cdot \text{SEM}$  wherein  $\text{SEM} = \text{SD} \cdot \sqrt{1 - \text{ICC}}$ , where SD was the pooled SD in the sample of which the ICC was determined [16, 17]. In addition, the MDC at group level was calculated by dividing the  $\text{MDC}_{\text{individual}}$  with the square root of  $n$ . To enable comparison of absolute reliability with other studies, the SEM and MDC were given as absolute values and as percentage of the mean. Heteroscedasticity was examined in a Bland-Altman plot [25]. The mean difference of the test and retest with limits of agreement was included in the plot.

### Responsiveness

The COSMIN panel recommends that responsiveness should be evaluated similarly as validity, i.e. by comparing changes on the instrument with changes on the gold standard. Since there is no gold standard in the measurement of PRO, responsiveness was determined by testing a hypothesis about expected correlations with changes, ESs and an SRM in the FJS and another questionnaire measuring the same construct. de Vet *et al.* stated that ES is appropriate when comparing different instruments [16] and SRM provides information on the magnitude of change in standardized units relative to variability of change [17].

When the assumption of normal distribution was fulfilled, Pearson's correlation coefficient was calculated between change scores of the FJS-12 and the HAGOS subscales. A comparison of the ES and the SRM between the FJS-12 and the subscales HAGOS-pain and -ADL were calculated, where ES was calculated as difference of mean/SD of the pre-operative score and the SRM was calculated as difference of mean/SD of the change score.

### Floor and ceiling effect

Distribution of the patients' scores (%) preoperative and at 6-month follow-up was calculated and floor and ceiling effects were defined to be present if more than 15% of the patients were reporting worst (0) or best (100) possible score. Terwee *et al.* [15].

**Table I. Demographics of patients included in the study of test-retest reliability and the study of responsiveness**

|                                  | Study of<br>test-retest<br>reliability (n = 50) | Study of<br>responsiveness<br>(n = 34) |
|----------------------------------|-------------------------------------------------|----------------------------------------|
| Women                            | 27 (54%)                                        | 18 (53%)                               |
| Age (years $\pm$ SD)             | 36 $\pm$ 19                                     | 36 $\pm$ 9                             |
| Height (cm $\pm$ SD)             | 175 $\pm$ 9                                     | 175 $\pm$ 9                            |
| Weight (kg $\pm$ SD)             | 78 $\pm$ 14                                     | 81 $\pm$ 16                            |
| BMI (kg/m <sup>2</sup> $\pm$ SD) | 25 $\pm$ 4                                      | 26 $\pm$ 5                             |

Current pathology, femoroacetabular impingement; procedure performed, hip arthroscopic treatment; SD, standard deviation.

## RESULTS

### Data

Of the 63 patients invited to the study of test-retest reliability three did not reply, two did not want to participate and seven reported a change in hip function of more than 20 points, meaning that they had to be excluded from the test-retest reliability analysis. Furthermore, one patient had an unlikely score of 0 out of 100 in the first test and 98 out of 100 in the retest 3 weeks later. This unusual score was not found in the patient's scores of the HAGOS, and thus it was decided to exclude the patient from the analysis.

For the study of the responsiveness all patients completed both questionnaires (Fig. 1).

For both the reliability and the responsiveness study, 3–4 reminder emails were sent if a patient did not complete the questionnaire within 3 days.

In the analysis of floor and ceiling effect, all questionnaires collected preoperative ( $n = 34$ ) and at 6-month follow-up ( $n = 58$ ) were included except for the patient with the unlikely score in the FJS-12.

### Patient characteristics

The characteristics of the patients included in the study of test-retest reliability and of the responsiveness were almost identical (Table I).

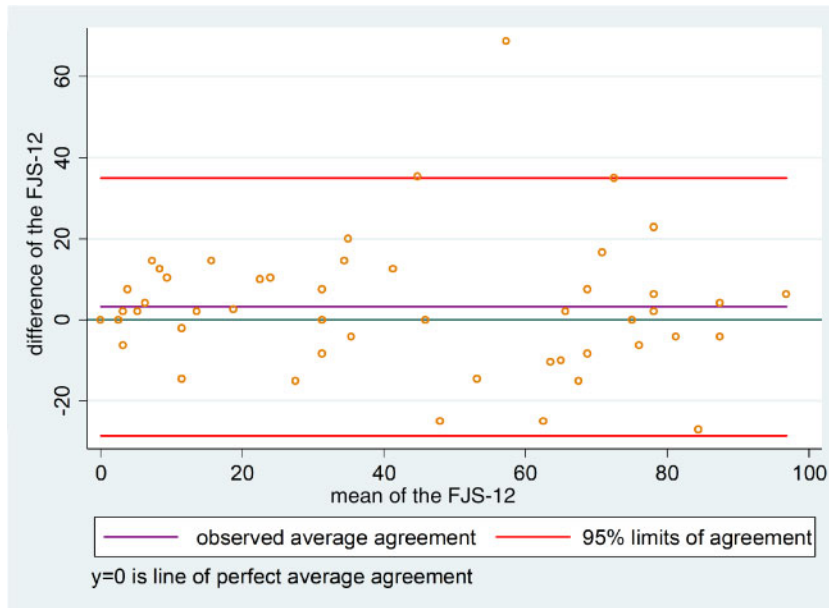
### Test-retest reliability

The FJS-12 was completed at 7(6–11) months postoperative and the interval between test and retest was 24(16–49) days. There was no significant difference between mean scores from test and retest of the FJS-12 (Table II). The test-retest assessment showed a high relatively reliability with an ICC of 0.9 for the FJS-12. The calculated

**Table II. The post-operative test–retest reliability of the FJS**

| n  | 1st FJS-12<br>mean (SD) | 2nd FJS-12<br>mean (SD) | Change score<br>(95% CI) | ICC<br>(95% CI) | SEM abs<br>(%) | MDC <sub>individual</sub><br>abs (%) | MDC <sub>group</sub><br>abs (%) |
|----|-------------------------|-------------------------|--------------------------|-----------------|----------------|--------------------------------------|---------------------------------|
| 50 | 46 (31)                 | 43 (32)                 | 3.2 (–1; 8)              | 0.9 (0.8; 0.9)  | 11 (26)        | 32 (72)                              | 4.5 (10)                        |

ICC, intraclass correlation coefficient; SEM, standard error of measurement; SEM (%), SEM in per cent of the mean of two test sessions; MDC, minimal detectable change (individual and group level); MDC (%), MDC in per cent of the mean of two test sessions; SD, standard deviation.



**Fig. 2.** Bland–Altman plot with limits of agreement showing difference between test and retest with the FJS-12 in patients following hip arthroscopic treatment. Mean difference between test–retest (purple line) and limits of agreement (red lines).

SEM was 11 and the MDC was 32 indicating the measurement error on an individual level and 4.5 on a group level.

The Bland–Altman plot with 95% limits of agreement (Fig. 2) for the results of the test–retest did not show signs of heteroscedasticity.

### Responsiveness

Responsiveness was evaluated by comparing the FJS-12 with the HAGOS-pain and the HAGOS-ADL from pre-operative to 6 months postoperatively (Table III). The FJS demonstrated a moderate level of responsiveness. We found a high correlation of 0.7 between change scores of the FJS-12 and the HAGOS-pain and a moderate correlation of 0.6 between the FJS-12 and the HAGOS-ADL. The ES for the FJS-12, the HAGOS-pain and HAGOS-ADL were almost the same, while the SRM for the FJS-12 had the lowest value compared with the two HAGOS-scales.

### Floor and ceiling effect

Since the frequency of maximum or minimum scores did not exceed 15%, there was no floor or ceiling effect of the FJS-12, the HAGOS-pain or the HAGOS-ADL.

## DISCUSSION

### Main findings

The FJS demonstrated high test–retest reliability. The level of responsiveness was high, and no relevant floor or ceiling effect was detected. This makes the FJS-12 very useful when evaluating the effect of FAIS.

### Test–retest reliability

The ICC found in this study (ICC = 0.9) was similar to the findings of the FJS in a study of Danish patients with knee joint replacement (ICC = 0.9) and similar to other PROs evaluated in patients with conservative or non-treated FAIS with ICC ranging from 0.7 to 0.9 [modified

**Table III. Analysis of the responsiveness of the FJS-12 from before surgery to 6-month follow-up**

|            | Baseline mean<br>(SD) n = 34 | 6 months mean<br>(SD) n = 34 | Change score<br>(SD) | Pearson's correlation<br>coefficient of change scores | ES  | SRM |
|------------|------------------------------|------------------------------|----------------------|-------------------------------------------------------|-----|-----|
| FJS        | 28 (23)                      | 43 (34)                      | 14 (28)*             |                                                       | 0.6 | 0.5 |
| HAGOS-pain | 55 (20)                      | 69 (25)                      | 14 (16)              | 0.7*                                                  | 0.7 | 0.8 |
| HAGOS-ADL  | 56 (26)                      | 69 (26)                      | 13 (20)*             | 0.6*                                                  | 0.5 | 0.7 |

\*Significant difference;  $P < 0.05$ . SD, standard deviation; ES, effect size; SRM, standardized response mean.

Harris Hip Score (mHHS): 0.76, Hip dysfunction and Osteoarthritis Outcome Score (HOOS): 0.84–0.96, Hip Outcome Score (HOS): 0.73–0.90, Non-Arthritic Hip Score: 0.94, International Hip Outcome Tool (iHOT): 0.93 and HAGOS: 0.79–0.94 [7, 26].

Furthermore, the ICC equalled findings in a study of the FJS with similar follow-up assessed in patients with knee joint replacements (ICC = 0.8) and in patients with hip joint replacements 16 months postoperative (ICC = 0.9) [6, 9].

The results of the  $MDC_{\text{individual}}$  (32) and the  $MDC_{\text{group}}$  (4.5) equalized the findings by Hinman *et al.* assessing the test-retest reliability of six different PROs among patients with FAIS ( $MDC_{\text{individual}} = 12.4\text{--}35.6$  and  $MDC_{\text{group}} = 2.2\text{--}7.3$ ) [26].

The measurement error at individual level was higher than findings in a Danish study investigating the MDC of the FJS in patients undergoing total knee replacement (MDC = 24) and higher than in patients with hip joint replacements (MDC = 21) [7, 9].

This high MDC value indicates a poor absolute reliability of the FJS-12 when applied in patients with FAIS. A clinical explanation of the wide dispersion of scores may be the very different levels of awareness and function among patients with FAIS during the rehabilitation period after hip arthroscopic treatment. The MDC value indicated that substantial change of more than 32 would be required to detect a change beyond measurement error at the individual level. In a clinical setting, the MDC can be regarded as the threshold to identify a statistically significant change in an individual patient and to determine whether an individual patient has made significant improvement. For research purposes the improvements after an intervention will be measured on group level rather than on the individual patient level. We suggest that the FJS-12 should be used to compare patients with FAIS on group level rather than on the individual patient level. The  $MDC_{\text{group}}$  calculated in this study, was acceptable as differences above 4.5 indicated a true change in the present group of 50 patients.

### Responsiveness

In line with our hypothesis, the responsiveness of FJS-12 was high (ES = 0.6, SRM = 0.5). A larger ES (3.6) and SRM (1.6) were found in patients with hip joint replacement assessed from preoperative to 1 year postoperative [4]. No other studies of the FJS-12 have investigated the ES and SRM from pre- to postoperative. However, our results of the ES were higher than findings in five other PROs (mHHS, HOOS, HOS, iHOT, HAGOS) in patients undergoing hip arthroscopic treatment on average 19 months earlier (ES = -0.1; 0.4) and equalled findings by Giesinger *et al.* (ES = 1.2, SRM = 0.8) assessing the postoperative responsiveness between 2 and 6 months of the FJS-12 in patients with knee joint replacement among four measurement tools (different from these of Kemp *et al.*) [5, 27].

It might be argued that the hip function level in patients with FAIS will continue to change after 6 months and that a 1-year follow-up could show larger change scores, ES and SRM [28]. However, a recent systematic review showed that the largest improvements in pain and ADL function happen during the first 6 months after surgery supporting our choice of a 6-month follow-up [21].

The FJS-12 change score showed a large variation in treatment effect contributing to almost similar values of the SRM among the two questionnaires. The high ES of the FJS-12 compared with the HAGOS subscales is explained by the size of the change scores. The choice of calculating the ES (ES or SRM) has a major influence on the results and further interpretation. Because the denominator in the calculation of the SRM examines variance of change instead of variance of baseline score, Katz *et al.* considered the SRM approach to be more informative [29].

### Floor and ceiling effect

There was no relevant floor or ceiling effect of either the FJS-12 or the HAGOS which has also been confirmed by several studies of the FJS-12 in patients with osteoarthritis

of the knee, in patients with knee joint replacement and in patients with hip joint replacement [3, 6, 8, 9, 11]. Preoperatively, it is seen as a strength of the FJS-12 that no patients obtained the maximum score and a limitation that four patients obtained the minimum score.

### Limitations

Regarding the study of responsiveness with 34 included patients, the ES and SRM might be overestimated as smaller studies tend to show larger treatment effects, because any set of small studies will likely contain a disproportionate number of very positive (or very negative) ESs [29]. This suggests that more patients should have been included. We did not calculate a sample size but merely followed the recommendation from de Vet *et al.* [16] and included 50 patients for the test–retest reliability study.

In addition, a sample size of minimum 50 patients is recommended by Terwee *et al.* [15].

The period for test–retest reliability seemed to be influenced by change of the hip function as there were seven patients (out of 58) with an unstable condition. This means that the test–retest reliability would possibly be higher if it had been investigated within a shorter time frame between assessments like other studies of FAIS, or later in the rehabilitation period in line with other PROs [26, 27]. We allowed a change of <20% on a VAS rating of the hip symptoms to represent a stable group. The results identified a very large MDC for interpretation on an individual level which indicates that the threshold of <20% change was too high to represent a group with stable hip symptoms.

### CONCLUSION

This study shows that FJS-12, a PRO with a different approach, namely assessing awareness of the hip joint, can be used as a supplement to other PROs. The FJS-12 demonstrated high relative reliability, while the values of the MDC in individuals and in groups were relatively high in post-operative patients following hip arthroscopic treatment for FAIS. The level of responsiveness to change was high and no relevant floor or ceiling effect was detected. This suggests that the FJS-12 might be an appropriate measurement tool in research for assessing awareness of the hip joint in young and physically active groups of patients undergoing hip arthroscopic treatment for FAIS.

### CONFLICT OF INTEREST STATEMENT

None declared.

### REFERENCES

1. U.S. Department of Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006; **4**: 1–2.
2. Gut R, Fuglsang M, Rimdal B. Spørg brugerne – en guide til kvalitative og kvantitative brugerundersøgelser i sundhedsvæsenet. *Enheden for Brugerundersøgelser, Region Hovedstaden* 2011; 6–60.
3. Behrend H, Giesinger K, Giesinger J *et al.* The “forgotten joint” as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty* 2012; **27**: 430–6.
4. Thienpont E, Vanden Berghe A, Schwab PE *et al.* Joint awareness in osteoarthritis of the hip and knee evaluated with the ‘forgotten joint’ score before and after joint replacement. *Knee Surg Sports Traumatol Arthrosc* 2016; **24**: 3346–51.
5. Giesinger K, Hamilton D, Jost B *et al.* Comparative responsiveness of outcome measures for total knee arthroplasty. *Osteoarthritis Cartilage* 2014; **22**: 184–9.
6. Baumann F, Ernstberger T, Loibl M *et al.* Validation of the German forgotten joint score (G-FJS) according to the COSMIN checklist: does a reduction in joint awareness indicate clinical improvement after arthroplasty of the knee? *Arch Orthop Trauma Surg* 2016; **136**: 257–64.
7. Thomsen M, Latifi R, Kallemose T *et al.* Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty. *Acta Orthop* 2016; **87**: 280–5.
8. Hamilton DF, Giesinger JM, MacDonald DJ *et al.* Responsiveness and ceiling effects of the forgotten joint score-12 following total hip arthroplasty. *Bone Joint Res* 2016; **5**: 87–91.
9. Shadid M, Vinken N, Marting L *et al.* The Dutch version of the forgotten joint score: test–retesting reliability and validation. *Acta Orthop Belg* 2016; **82**: 112–8.
10. Matsumoto M, Baba T, Homma Y *et al.* Validation study of the forgotten joint score-12 as a universal patient-reported outcome measure. *Eur J Orthop Surg Traumatol* 2015; **25**: 1141–5.
11. Thompson S, Salmon L, Webb J *et al.* Construct validity and test re-test reliability of the forgotten joint score. *J Arthroplasty* 2015; **30**: 1902–5.
12. 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**: 1169–76.
13. Ganz R, Parvizi J, Beck M *et al.* Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res* Dec 2003; **417**: 112–20.
14. Thorborg K, Hölmich P, Christensen R *et al.* The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med* 2011; **45**: 478–91.
15. Terwee CB, Bot SDM, de Boer MR *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60**: 34–42.
16. de Vet HCW, Terwee CB, Mokkink LB *et al.* *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press, 2011.

17. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*, 3rd edn. Upper Saddle River, NJ: Pearson/Prentice Hall, 2009.
18. Mokkink L, Terwee C, Knol D *et al.* The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; **10**: 22.
19. Kierkegaard S, Lund B, Dalgas U *et al.* The Horsens–Aarhus Femoro Acetabular Impingement (HAFAI) cohort: outcome of arthroscopic treatment for femoroacetabular impingement. Protocol for a prospective cohort study. *BMJ Open* 2015; **5**: e008952.
20. Mokkink L, Terwee C, Patrick D *et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; **19**: 539–49.
21. Kierkegaard S, Langeskov-Christensen M, Lund B *et al.* Pain, activities of daily living and sport function at different time points after hip arthroscopy in patients with femoroacetabular impingement: a systematic review with meta-analysis. *Br J Sports Med* 2017; **51**: 572–9.
22. Wild D, Grove A, Martin M *et al.* Principles of Good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005; **8**: 94–104.
23. Thomeé R, Jónasson P, Thorborg K *et al.* Cross-cultural adaptation to Swedish and validation of the Copenhagen Hip and Groin Outcome Score (HAGOS) for pain, symptoms and physical function in patients with hip and groin disability due to femoroacetabular impingement. *Knee Surg Sports Traumatol Arthrosc* 2014; **22**: 835–42.
24. World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013; **310**: 2191–4.
25. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; **26**: 217–38.
26. Hinman RS, Dobson F, Takla A *et al.* Which is the most useful patient-reported outcome in femoroacetabular impingement? Test–retest reliability of six questionnaires. *Br J Sports Med* 2014; **48**: 458–63.
27. Kemp JL, Collins NJ, Roos EM *et al.* Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 2013; **41**: 2065–73.
28. Fairley J, Wang Y, Teichtahl AJ *et al.* Management options for femoroacetabular impingement: a systematic review of symptom and structural outcomes. *Osteoarthritis Cartilage* 2016; **24**: 1682–96.
29. Katz JN, Larson M, Phillips C *et al.* Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; **30**: 917–25.

## APPENDIX 1: The Danish version of the Forgotten Joint Score

### Hofte-spørgeskema (Score for glemte led – 12)

Patient: \_\_\_\_\_ Dato: \_\_\_\_\_

Et sundt led er ikke noget man er opmærksom på i hverdagen. Selv de mindste problemer kan imidlertid gøre en opmærksom på et led. Dette betyder, at du tænker på dit led, eller at din opmærksomhed rettes mod leddet. De følgende spørgsmål vedrører, hvor ofte du er opmærksom på det berørte hofteled i hverdagen.

Du bedes vælge det mest relevante svar på hvert spørgsmål.

| Er du opmærksom på dit hofteled...                      | Aldrig                | Næsten aldrig         | Sjældent              | Nogle gange           | For det meste         |
|---------------------------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1. ... i sengen om natten?                              | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2. ... når du sidder på en stol i mere end én time?     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3. ... når du går i mere end 15 minutter?               | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4. ... når du tager et bad/brusebad?                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5. ... når du kører i bil?                              | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 6. ... når du går op ad trapper?                        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 7. ... når du går i ujævnt terræn?                      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 8. ... når du rejser dig fra en lavt siddende stilling? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 9. ... når du står op i længere tid?                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 10. ... når du udfører hus- eller havearbejde?          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 11. ... når du går en tur/er på vandretur?              | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 12. ... når du dyrker din yndlings sport?               | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |