The Victorian Institute of Sport Assessment Scale for Patellar Tendinopathy (VISA-P): A Reliability Generalization Meta-analysis

Antonio Palazón-Bru, PhD,* María Isabel Tomás Rodríguez, PhD,† Emma Mares-García, PhD,* Sergio Hernández-Sánchez, PhD,† María Ángeles Carbonell-Torregrosa, MD, PhD,* and Vicente Francisco Gil-Guillén, MD, PhD*

Abstract

Objective: The Victorian Institute of Sport Assessment-Patella (VISA-P) is a questionnaire to assess the severity of patellar tendinopathies. Its use requires good reliability indicators: internal consistency, test-retest and parallel forms. Several studies have been published examining this question, but to date the reliability of this questionnaire (meta-analysis) has not been generalized. The aim of this study was to perform a meta-analysis to generalize the reliability of the VISA-P. **Data sources:** MEDLINE, EMBASE, and Scopus. **Study selection:** Studies included were those examining the reliability coefficients of the VISA-P: Cronbach alpha, intraclass correlation coefficient (ICC), and parallel-forms (correlation coefficients compared with other scales). **Data extraction:** All coefficients were extracted and the mean reliability was obtained using fixed- or random-effects models. Sensitivity (leave-one-out analysis) was analyzed. Quality assessment was performed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. **Data Synthesis:** Of 364 scientific articles, 12 fulfilled meta-analysis criteria. The summary statistic was 0.86 [95% confidence interval (CI): 0.78-0.92] for Cronbach alpha and 0.94 (95% CI: 0.89-0.97) for the ICC. Parallel forms depended on the comparative test used, ranging from −0.83 to 0.68. The sensitivity analysis found an influential study for the parallel-forms reliability in the Blazina score. We were unable to analyze the asymmetry of funnel plots and meta-regression models because of the number of studies. **Conclusions:** The reliability of VISA-P for assessing the severity of patellar tendinopathies requires greater evaluation with more scientific evidence before it can be implemented in clinical practice.

Key Words: VISA-P, reliability generalization, meta-analysis, knee injuries, sports physiotherapy

(*Clin J Sport Med* 2020;00:1–10)

INTRODUCTION

Patellar tendinopathies are common injuries in sport practitioners, especially in jumping or sprinting disciplines. The Victorian Institute of Sport Assessment-Patella (VISA-P) is the most widely used condition-specific patient-reported questionnaire to assess the clinical impact of this injury. It was developed originally in English to evaluate the severity of symptoms in patellar tendinopathies and the repercussions of these on daily and physical activities, through 8 questions. Six of these assess the pain in different activities or positions on a scale of 0-10, with 0 being no pain and 10 being the maximum possible pain, and 2 items provide information on the sports practice of the subject through categorized responses. The maximum score on the VISA-P scale is 100, which represents an asymptomatic subject, whereas the minimum score (0 points) corresponds to total functional impairment.[1]

The standard error of measurement of the VISA-P ranges from 4 to 5.4 points and its minimum detectable change thresholds have been established between 11 and 12.2 points, respectively.[2,3] The minimum clinically important change threshold estimated for the VISA-P scale in athletes with chronic patellar tendinopathy is an average change greater than 13 points, although this depends on the baseline score, and also on the interpretation of minimum clinically significant or relevant change on the scale of perceived global change.[2] Other authors have noted that the VISA-P is not sensitive to very small changes in this tendinopathy and considering its slow clinical progress recommend that the VISA-P should be used at intervals of 4 weeks or more.[4] The correlation between the VISA-P score and the extent of tissue pathology has been reported, noting that the VISA-P scale scoring reflects the extent of tendon tissue pathology.[5]

In the return-to-sport process in tendinopathy, pain/symptoms, in addition to other factors, need to be considered.[6] The use of the VISA-P scale in this context has become helpful in clinical decision-making. For example, the Medical Staff of FC Barcelona, in their guide to clinical practice for tendinopathies, specifically incorporate the evolution of the VISA-P scores to determine return-to-play in soccer players with patellar tendinopathy.[7] Concretely, the VISA-P score requirements to be declared medically fit are: the soccer player

Submitted for publication August 12, 2019; accepted October 26, 2019.

*From the *Department of Clinical Medicine, Miguel Hernández University, San Juan de Alicante, Alicante, Spain; and †Department of Pathology and Surgery, Center for Translational Research in Physiotherapy, Miguel Hernández University, San Juan de Alicante, Alicante, Spain.*

The authors report no conflicts of interest.

M. Á. Carbonell-Torregrosa and V. F. Gil-Guillén contributed equally.

Corresponding Author: Antonio Palazón-Bru, PhD, Clinical Medicine Department, Miguel Hernández University, Carretera de Valencia, Alicante S/N, 03550 San Juan de Alicante, Spain (antonio.pb23@gmail.com).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.cjsportmed.com).

http://dx.doi.org/10.1097/JSM.0000000000000810

with a score lower than 50 should not return to work on the field; the player must have a score of over 60 before rejoining the rest of the group, and there must have been an improvement of at least 30 points since the first time the questionnaire was completed, as long as the overall total is more than 60 points.

More recently, the VISA-P score has been included in a proposed patellar tendinopathy screening tool, which aims to identify risk factors for patellar tendinopathy considering together intrinsic and extrinsic risk factors identified in the literature and the score of 2 subjective outcome measures, such as the visual analogue and VISA-P scales.[8]

For the direct use of the VISA-P in a given population, as in all questionnaires, it must meet reliability standards, which include the analysis of internal consistency, test-retest, and parallel and interobserver forms.[9] Note that in the VISA-P the latter does not apply, because it is the patient alone who can answer the VISA-P questions.[1]

Since the original publication of the questionnaire, it has been translated into several languages to determine its reliability in them,[10–12] following the procedure to adapt the questionnaire to the cultural region where it is to be used.[13] The process should aim to produce the equivalent of the original scale, adapting it in a culturally relevant and understandable manner, while maintaining the meaning and intent of the original article.[14] However, we must bear in mind that reliability depends on the scores obtained in the test in a specific application of the test in a specific group of subjects.[15] In other words, the coefficients that measure reliability may vary in the different applications of the questionnaire being evaluated. However, it is very common for researchers to assume the reliability of a questionnaire in previous applications of the test in other samples of subjects,[16] which is known as induction of reliability.[15] To minimize this type of error, the best method is to perform a meta-analysis to generalize the reliability of a given questionnaire, considering all the results already published on the reliability of the questionnaire to be evaluated.[17]

Taking into account the above considerations and that to the best of our knowledge no meta-analysis has been performed to generalize the reliability of the VISA-P, we conducted a study using the MEDLINE, EMBASE, and Scopus databases to determine the mean reliability of the VISA-P. In addition, we investigated which characteristics of the studies may affect reliability.

MATERIAL AND METHODS

The PRISMA statement has been followed for reporting a systematic review with meta-analysis.[18]

Protocol

This review has been registered in the PROSPERO database (CRD42019127782).

Search Strategy

This is a systematic review of articles indexed in the MEDLINE (through PubMed) and EMBASE databases from inception until December 31, 2018 that analyzed the reliability of the VISA-P questionnaire, either in its original version or in translations into other languages. In other words, we only included validation studies of the VISA-P. The keyword used was "VISA-P." Furthermore, we filtered the results by language (English and Spanish), selecting those papers that had an abstract. Moreover, the citations of the original publication by Visentini were analyzed through Scopus,[1] and the references of the extracted papers.

Inclusion Criteria

For the systematic review, we selected scientific articles that analyzed the reliability of the VISA-P with original data from patients with patellar tendinopathy. For the meta-analysis, we only included those reporting coefficients appropriate for the analysis (at least one coefficient), according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria[19]: (1) Internal consistency (Cronbach alpha), (2) Test-retest (intraclass correlation coefficient, ICC), and (3) Parallel forms (correlation coefficients comparing with other scales). We assessed parallel-forms by extracting correlations between VISA-P and other related scale(s). Although in the COSMIN statement this has been defined as convergent validity,[19] in the reliability generalization literature it is generally referred to as "parallel forms."

Exclusion Criteria

Articles that were only an abstract, a review, or a letter to the editor were excluded. For the meta-analysis, independently for each component, we excluded those papers with an inadequate design for the analysis of internal consistency, test-retest, and parallel forms (see Quality assessment).[19]

Study Selection and Data Extraction

Screening of titles and abstracts was done in a paired and blinded fashion by 2 reviewers (A.P.-B. and M.I.T.R.) and was based on the inclusion and exclusion criteria to determine the potential articles to be comprehensively analyzed. In the event of disagreement, a consensus was reached with a third researcher (V.F.G.-G.), although there was always agreement. This same process was followed when the full text of each article was assessed. Finally, the references of the studies included in the meta-analysis were also revised.

From the selected articles, the following information was extracted: subjects; percentage of male gender; average age (years); country; sample size; and internal consistency (Cronbach alpha coefficient for the total score), test-retest (ICC with the time interval), and parallel forms (correlation coefficients with similar scales) coefficients. If more than one measurement for Cronbach alpha coefficient was submitted, the result of the first application of the questionnaire was selected. If this situation occurred in the ICC (administration of the questionnaire 3 or more times), the value was selected by comparing the first administration with the second, unless this was done immediately, in which case we analyzed the ICC between the first and third administration. Finally, only statistically significant correlation coefficients, ie, those that showed a relationship with the VISA-P, were taken into account to analyze the parallel forms reliability.

Quality Assessment

We assessed the quality of the studies found using the COSMIN checklist.[19] This tool determines the quality of different aspects of a patient-reported outcomes measure, including validity and reliability. Because the aim of this review was to determine the reliability of the VISA-P, the COSMIN sections on the reliability of internal consistency (Box 4), test-retest (Box 6), and parallel forms (Box 9a) were analyzed.[19] Each box contains a series of items that can take the following values: very good, adequate, doubtful, or inadequate. To assess the overall quality of each of the sections, the "worst score counts" principle was used. This is defined as the worst score of the items.[19]

Data Analysis

Different analyses were performed for the 3 types of coefficients that analyze the reliability of the questionnaire. In addition, to approximate the distribution of these coefficients to a normal distribution and to stabilize the variances, the Hakstian and Whalen transformation was applied to Cronbach alpha, and Fisher conversion to the ICC and to the correlation coefficients (parallel forms).[20,21] To obtain global estimates of reliability coefficients, fixed-effects models were used when there was no heterogeneity between the studies, which was assessed as having an I^2 statistic value greater than 50% and/or a P-value <0.10 associated with Cochran's Q statistic. Otherwise, the coefficients were estimated through mixed-effects models.[22] Note that when 2 studies were available, a fixed-effect model was applied, because there were not enough data available to estimate heterogeneity.[23] The method used to estimate the between-study variance was restricted maximum likelihood.

Potential sources of heterogeneity of results were assessed using subgroup analyses: study quality (very good, adequate and doubtful) and continent (Europe, America, Asia, Oceania and Africa). No groupings were made over time of the successive applications, because it was an item evaluated in COSMIN, within the design.[19]

To determine the influence of studies included in the meta-analysis (sensitivity analysis), the leave-one-out method was used, which is equivalent to repeating the calculations excluding a single study each time and comparing the results obtained. A study was considered influential if it varied the overall coefficient by at least 10%. We assessed the asymmetry of funnel plots and the Egger test when we obtained at least 10 studies in the analyzed parameter.[24]

Type I error was set at 5% and for each relevant parameter its associated confidence interval (CI) was calculated. All statistical calculations were performed with R 3.3.3 through the metafor package (Meta-Analysis Package for R).

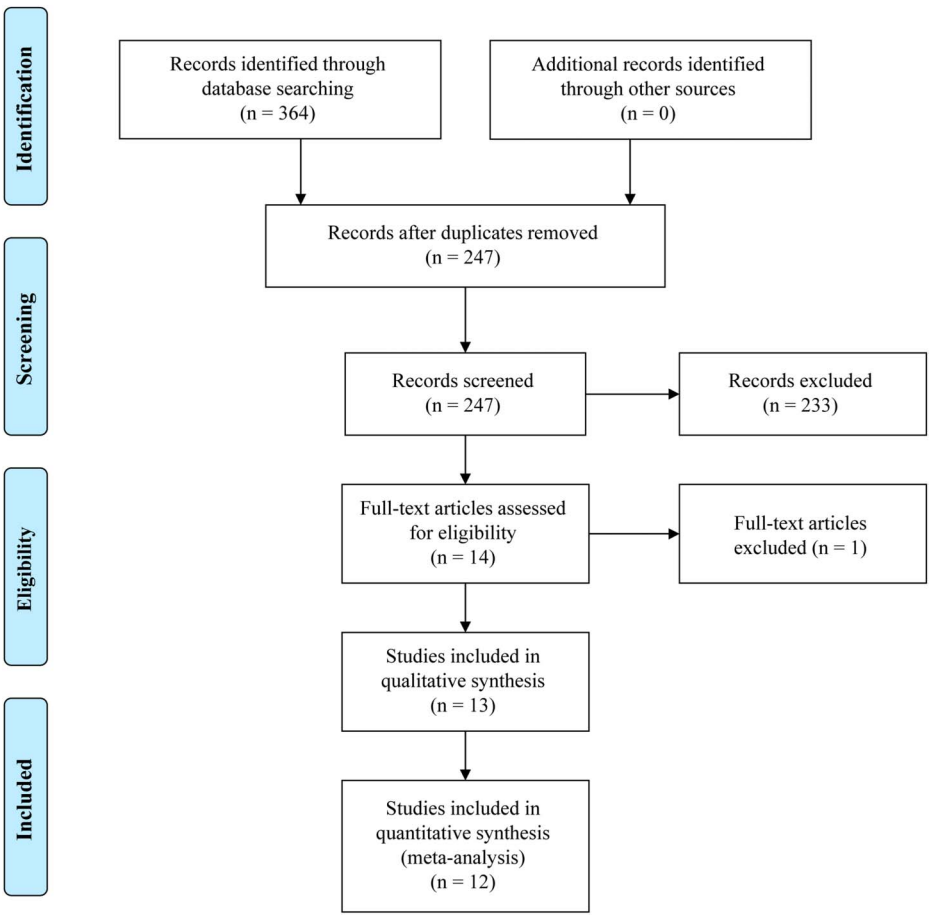

Figure 1. Flow chart of the eligible papers used in the meta-analysis (PRISMA statement).

TABLE 1. Summary of the Main Data From the Studies Included in the Systematic Review

Study	Subjects	Males (%)	Mean Age (yrs)	Country	Sample Size	Cronbach Alpha	ICC	Time Interval	Parallel Forms (S/P)
Visentini et al[1]	Normal volunteers	Not given	31	Australia	26	N/C	N/C	1 wk	−0.93 (P) Nirschl score
Visentini et al[1]	Sports clinic patients without jumper's knee	Not given	27	Australia	26	N/C	N/C	1 wk	−0.47 (P) Nirschl score
Visentini et al[1]	Elite basketball players	Not given	24	Australia	100	N/C	N/C	1 wk	N/C
Visentini et al[1]	Jumper's knee patients (clinic)	Not given	25	Australia	14	N/C	N/C	1 wk	−0.74 (P) Nirschl score
Visentini et al[1]	Presurgical patients for chronic jumper's knee	Not given	31	Australia	15	N/C	N/C	1 wk	−0.76 (P) Nirschl score
Visentini et al[1]	Postsurgical patients (6 months) for chronic jumper's knee	Not given	32	Australia	15	N/C	N/C	1 wk	−0.71 (P) Nirschl score
Visentini et al[1]	Postsurgical patients (12 months) for chronic jumper's knee	Not given	32	Australia	15	N/C	N/C	1 wk	−0.69 (P) Nirschl score
Frohm et al [25]	Healthy students (n = 17), male basketballers (n = 17), non-surgically treated patients with patellar tendinopathy (n = 17)	82.4	24	Sweden	51	0.83	0.97	4-7 d	N/C
Maffulli et al [26]	Patellar tendinopathy	100	27.9	Italy	25	N/C	N/C	30 min	N/C
Zwerver et al[27]	Healthy students (n = 18), volleyballers (n = 15), patellar tendinopathy (n = 14), patients who had surgery for patellar tendinopathy 6 months before (n = 6), patients who had other knee injuries (n = 17), and patients with symptoms unrelated to their knees (n = 19).	25	19.2	Netherlands	89/71*	0.73	0.74	2.5 wk	N/C
Lohrer and Nauck[11]	Patellar tendinopathy with conservative treatment	Not given	34.8	Germany	23	0.876	0.878	1 wk	N/C
Lohrer and Nauck[11]	Healthy individuals	Not given	29.8	Germany	57	N/C	0.872	1 wk	N/C
Lohrer and Nauck[11]	Healthy individuals (n = 57) and patellar tendinopathy with conservative treatment (n = 23)	Not given	31.2	Germany	80	N/C	N/C	1 wk	−0.81 (S) Blazina
Hernandez-Sánchez et al[10]	Healthy students (n = 40), athletes (n = 40), patellar tendinopathy (n = 40) and other knee injuries (n = 30)	86.7	23.5	Spain	150	0.885	0.994	7-10 d	−0.05 to 0.65 SF-36 0.897 Kujala 0.782 Cincinnati (S)
Park et al[28]	Volleyballers (with patellar tendinopathy, n = 23; without the diagnosis n = 5)	46.4	15.9	Korea	28	0.80	0.96	1 wk	N/C
Wageck et al[3]	Patellar tendinopathy	73.1	23.4	Brazil	52	0.76	0.91	24-48 h	0.60 Lysholm (P)
Korakakis et al[29]	Healthy individuals (n = 61), athletes (n = 34), patellar tendinopathy (n = 32), and other knee injuries (n = 30)	51.3	26.3	Greece	187	0.785	0.818	15-17 d	−0.839 Blazina (S)
Kaux et al[12]	Healthy individuals (n = 22), athletes (n = 42), and patellar tendinopathy (n = 28)	71.3	22.6	Belgium	92	N/C†	0.99	30 min	0.16 to 0.72 SF-36 (S)
Celebi et al[30]	Healthy individuals (n = 31), volleyballers (n = 24), and patellar tendinopathy (n = 34)	59.6	24.4	Turkey	89	0.79	0.96	24 h	0.473 VAS 0.419 participation to sports‡ (P)

TABLE 1. Summary of the Main Data From the Studies Included in the Systematic Review *(Continued)*

Study	Subjects	Males (%)	Mean Age (yrs)	Country	Sample Size	Cronbach Alpha	ICC	Time Interval	Parallel Forms (S/P)
Hernandez-Sanchez et al[31]	Athletes with patellar tendinopathy who had undergone physiotherapy treatment	59.4	29.1	Spain	249	0.74	N/C	N/C	N/C
Acharya et al[32]	Athletes with (n = 35) and without (n = 35) patellar tendinopathy	Not given	18.9	India	70	N/C	N/C	1 wk	0.72 (S) Blazina
Acharya et al[32]	Athletes with patellar tendinopathy	Not given	18.9	India	35	0.99	0.97	1 wk	N/C
Acharya et al[32]	Athletes without patellar tendinopathy	Not given	19.0	India	35	0.98	0.96	1 wk	N/C

** There were 18 patients who did not complete the second administration of the test. Therefore, the sample size for the ICC was 71.*
† Not calculated for the overall score of the questionnaire, but the authors estimated it for each item.
‡ This score was not defined and referenced in the paper.
N/A, not applicable; N/C, not calculated; P, Pearson correlation coefficient; S, Spearman correlation coefficient; SF-36, Short Form Health Survey (the authors analyzed the components of the questionnaire); VAS, visual analogue scale.

RESULTS

Figure 1 shows the flowchart of the systematic review. Combining the MEDLINE and EMBASE databases with the citation analysis of the original article by Visentini (Scopus) provided 364 scientific articles. After examining citations for duplicates, 117 were detected and excluded from the analysis of titles and abstracts, leaving a total of 247 articles to be screened. After examining these titles and abstracts, 14 papers were considered for inclusion in the qualitative synthesis [1,3,10–12,25–33]; that is, those in which the full text would be analyzed. Of these 14, one paper was eliminated both for the systematic review and the meta-analysis, because it was a congress abstract.[33] Moreover, another did not have the necessary reliability coefficients for this study. This was the study developed in Italy,[26] which assessed internal consistency with a Kappa coefficient instead of Cronbach alpha. Test-retest reliability was estimated through Pearson correlation coefficient and performed with a margin of only 30 minutes and did not determine parallel forms reliability. However, this study was assessed in the systematic review. Therefore, 12 articles were finally included in the meta-analysis and 13 for the systematic review. The main features of the 13 papers are shown in Table 1. Most of these studies included a control group, another group of people at risk of developing patellar tendinopathy, and a final group of patients with this condition. The countries of origin belonged to Europe, Asia, Oceania, and America, and the sample size ranged from 23 to 249 individuals. Note that many studies performed stratified analyses by groups and in Table 1, we have given the information in this manner. The percentage of men ranged from 25% to 100%, although there were 12 groups for whom this information was not available from 3 studies.[1,11,32] The mean age was indicated in all groups and papers, ranging from 15.9 to 34.8 years. Table 1 also shows all the reliability coefficients analyzed, indicating their value when they were estimated, the test-retest time and the type of correlation coefficient used for the parallel forms reliability (Pearson or Spearman). It should be noted that the test-retest time had great variability, with a minimum value of 30 minutes and a maximum of 2.5 weeks.

Table 2 details the quality assessment of the studies for the internal consistency, test-retest and parallel forms reliability for each COSMIN domain analyzed. Regarding internal consistency, it is noted that all studies were of very good quality, except for the original Visentini study and the Italian validation by Maffulli, which were inadequate and therefore excluded for the meta-analysis of this type of reliability, because they did not use Cronbach alpha coefficient for this purpose.[1,26] Second, we found that test-retest reliability was adequate in 6 of the studies analyzed (50%), doubtful in 4 (33.3%), and inadequate in the remaining 2 (16.7%), because of the very short period of time (30 minutes). Finally, parallel forms reliability was observed, analyzing each instrument of comparison with the VISA-P separately. When the authors obtained this coefficient, an adequate quality was seen, except in the *sports participation* of the Turkish research group, because it did not give any kind of detail or reference on this element of comparison.

Figure 2 shows the summary statistic for Cronbach alpha, obtained through a random-effects model, which had a value of 0.86 (95% CI: 0.78-0.92). The overall estimate of the ICC (Figure 3) was also obtained through a random-effects model and had a magnitude of 0.94 (95% CI: 0.89-0.97). Regarding the analysis of parallel forms reliability (Figure 4), because different comparative tests with the VISA-P were used, the analysis had to be done according to the chosen test. The first test was the Blazina score (Figure 4A), which is based on criteria of pain evolution according to functionality.[34] Through a Spearman correlation coefficient, using a random-effects model, an overall estimate of −0.44 (95% CI: −0.95 to 0.71) was obtained. Other tests applied were 3 components of the Short Form Health Survey (SF-36), a questionnaire in which a higher score implies better quality of life in each of the components[35]: physical function (Figure 4B), physical role (Figure 4C), and bodily pain (Figure 4D). Using a fixed-effects model, the first component obtained an estimate of 0.68 (95% CI: 0.61-0.74) in the Spearman correlation coefficient. The second component, using a fixed-effects model, yielded an estimate of 0.51 (95% CI: 0.41-0.60) in

TABLE 2. Summary of Quality Assessment for the Studies Included in the Systematic Review

	Internal Consistency			Test-Retest Reliability			Parallel Forms Reliability			
Study	Design	Statistical Methods	Overall	Design	Statistical Methods	Overall	Comparator	Design	Statistical Methods	Overall
Visentini et al[1]	Inadequate	Inadequate	Inadequate	Adequate	Adequate	Adequate	Nirschl score	Adequate	Adequate	Adequate
Frohm et al[25]	Very good	Very good	Very good	Adequate	Adequate	Adequate	N/A	Inadequate	Inadequate	Inadequate
Maffulli et al[26]	Very good	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	N/A	Inadequate	Inadequate	Inadequate
Zwerver et al[27]	Very good	Very good	Very good	Doubtful	Adequate	Doubtful	N/A	Inadequate	Inadequate	Inadequate
Lohrer et al[11]	Very good	Very good	Very good	Adequate	Very good	Adequate	Blazina	Very good	Adequate	Adequate
Hernandez-Sánchez et al[10]	Very good	Very good	Very good	Adequate	Adequate	Adequate	SF-36 Kujala Cincinnati	Adequate very good very good	Adequate	Adequate
Park et al[28]	Very good	Very good	Very good	Adequate	Adequate	Adequate	N/A	Inadequate	Inadequate	Inadequate
Wageck et al[3]	Very good	Very good	Very good	Doubtful	Very good	Doubtful	Lysholm	Very good	Adequate	Adequate
Korakakis et al[29]	Very good	Very good	Very good	Doubtful	Very good	Doubtful	Blazina	Very good	Adequate	Adequate
Kaux et al[12]	Very good	Very good	Very good	Inadequate	Very good	Inadequate	SF-36	Adequate	Very good	Adequate
Celebi et al[30]	Very good	Very good	Very good	Doubtful	Adequate	Doubtful	VAS SP	Very good inadequate	Adequate	Adequate inadequate
Hernandez-Sanchez et al[31]	Very good	Very good	Very good	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Acharya et al[32]	Very good	Very good	Very good	Adequate	Very good	Adequate	Blazina	Very good	Very good	Very good

The paper published by Hernandez-Sanchez et al in 2017, previously, analyzed test-retest and parallel-forms reliability in 2011.
SF-36, Short Form Health Survey (the authors analyzed the components of the questionnaire); SP, sports participation; VAS, visual analogue scale.

the coefficient. The last component, also using a fixed-effects model, obtained a summary statistic of 0.65 (95% CI: 0.57-0.72). Finally, in the study by Visentini, the Nirschl score was analyzed in different groups with a Pearson correlation coefficient, which allowed a meta-analysis of them (Figure 4E),[1] obtaining through a random effects model an overall estimate of −0.75 (95% CI: −0.87 to −0.56).

The subgroup analysis is shown in the **Supplemental Digital Content 1** (see **Appendix**, http://links.lww.com/JSM/A224), which stratified the results by continent and study quality according to the COSMIN statement. Only statistically significant differences were found in the ICC by continent ($P = 0.017$). These differences were established as the Asian studies obtained greater test-retest reliability than the rest of the continents. No differences were found for either Cronbach alpha or the quality analysis to detect possible sources of heterogeneity. For obvious reasons, this analysis could not be performed for parallel forms reliability, because either there were few studies in each comparator, they were all of the same quality, or they were performed on the same continent.

The sensitivity analysis, using the leave-one-out method, is reflected in the **Supplemental Digital Content 1** (see **Appendix**, http://links.lww.com/JSM/A224), where it is seen that none of the coefficients are altered in excess when a single study is eliminated, except for 2 components of the SF-36 (physical role and bodily pain), the Blazina classification system and the Nirschl score. The asymmetry of the funnel plots could only be assessed for Cronbach alpha and ICC, as a maximum of 9 studies were available (<10), showing no evidence of asymmetry (see **Appendix**, **Supplemental Digital Content 1**, http://links.lww.com/JSM/A224).

DISCUSSION

The purpose of this study was to determine the reliability of the VISA-P questionnaire for measuring the severity of symptoms in patellar tendinopathies[1] through a study with the design that provides the best scientific evidence (meta-analysis), known as reliability generalization.[21] With this analysis, overall coefficients were obtained to determine 3 aspects: internal consistency, test-retest reliability, and parallel forms. First, a Cronbach alpha value (internal consistency) of 0.86 was obtained. Furthermore, the ICC reached a summary value of 0.94. The standard thresholds for adequate reliability for patient-reported outcome measures for group (clinical trials) and individual (patients) comparisons have been suggested to be 0.7 and 0.9, respectively. For the VISA-P questionnaire, the results obtained for the reliability indicators reached these limits.[36] Below these standards, the use of the VISA-P scores obtained in the assessment of a patient would seriously affect its validity. For example, if there is an ICC below the standards, a change in the score could be assumed as real without it having occurred, with the negative consequences that this entails in the clinical management of the patient with patellar tendinopathy.

Finally, the magnitude (absolute value) of the parallel forms reliability varied, according to the comparative test, between 0.51 and 0.83. This, according to Cohen, is classified as a large correlation.[37] In short, if we analyze just the value of the coefficients (without taking into account sensitivity analysis, publication bias or heterogeneity), all the reliability parameters through the generalization of the results obtained in the literature were very satisfactory.

It should be noted that the results obtained in our meta-analysis are a consequence of the information from the

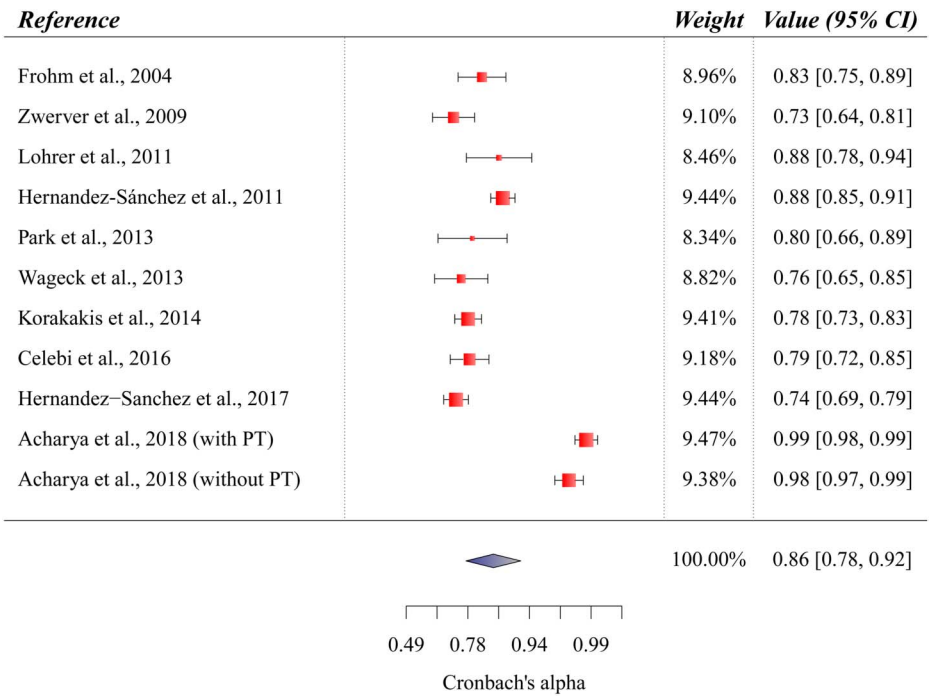

Figure 2. Forest plot of Cronbach alpha coefficient for the Victorian Institute of Sport Assessment scale-Patella (VISA-P). We used a random-effects model. Heterogeneity analysis: Q = 418.54, df = 10, $P < 0.001$; $I^2 = 96.56\%$.

scientific papers found in our review and, therefore, we were unable to examine certain aspects. First, parallel forms reliability was analyzed in only 5 studies by 5 different comparative tests,[1,10–12,29] such that little information was available to assess this aspect of reliability, including an analysis of the asymmetry of the funnel plots. Furthermore,

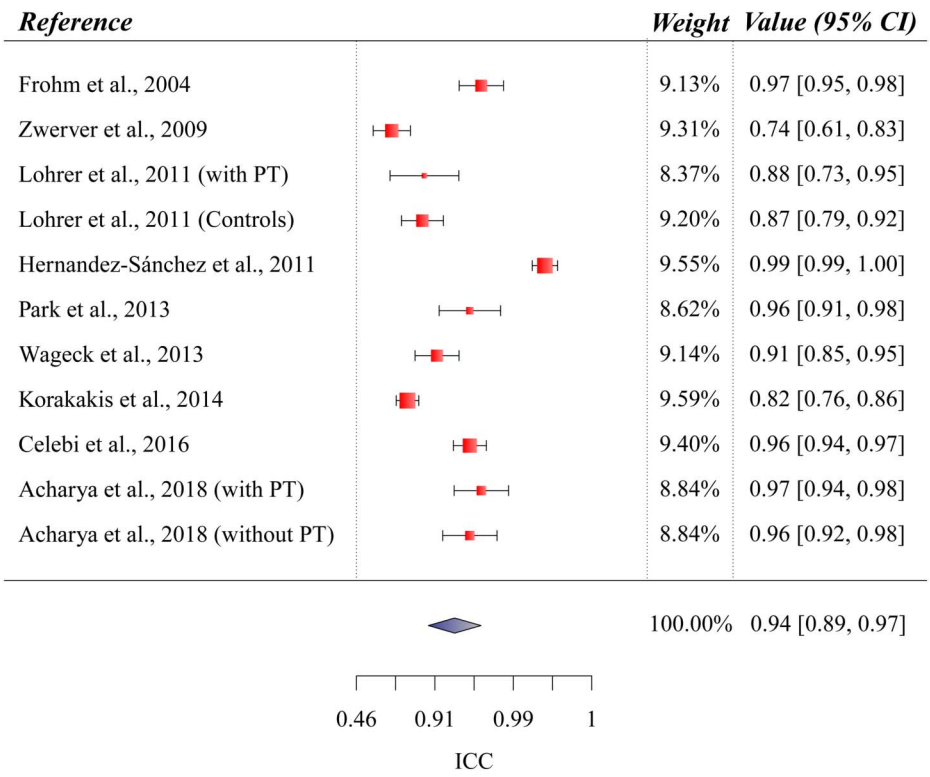

Figure 3. Forest plot of the ICC for the Victorian Institute of Sport Assessment scale-Patella (VISA-P). We used a random-effects model. Heterogeneity analysis: Q = 333.71, df = 10, $P < 0.001$; $I^2 = 95.05\%$. PT, patellar tendinopathy.

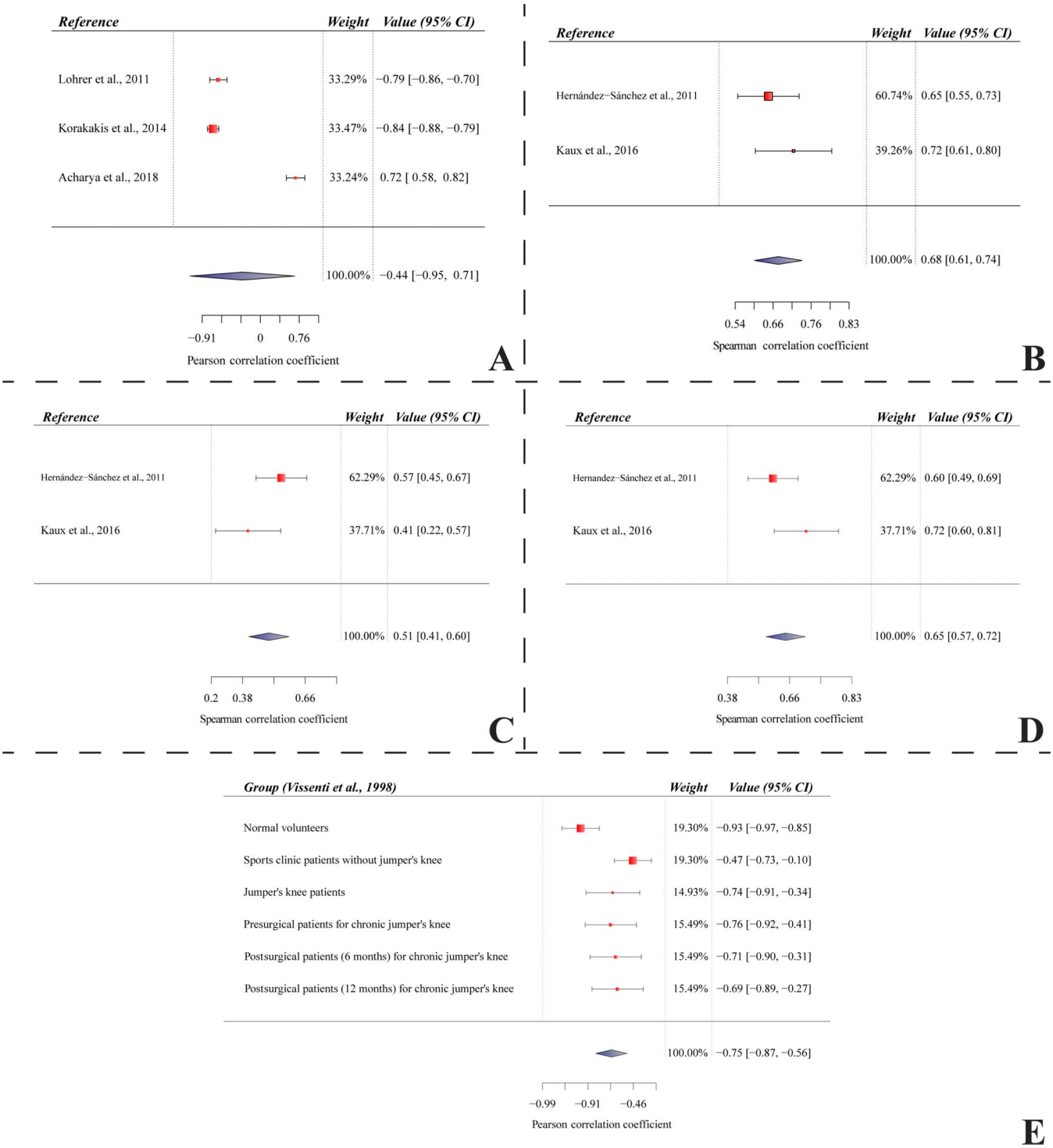

Figure 4. Forest plot of Pearson correlation coefficient for the Victorian Institute of Sport Assessment scale-Patella (VISA-P). A, Comparison with the Blazina score: We used a random-effects model. Heterogeneity analysis: Q = 232.86, df = 2, $P < 0.001$; $I^2 = 99.27\%$. B, Comparison with the Short Form Health Survey (SF-36) in the physical function: We used a fixed-effects model. C, Comparison with the SF-36 in the physical role: We used a fixed-effects model. D, Comparison with the SF-36 in bodily pain: We used a fixed-effects model. E, Comparison with the Nirschl score (only used by Vissentini et al, 1998): We used a random-effects model. Heterogeneity analysis: Q = 15.95, df = 5, $P = 0.007$; I 2 = 64.23%.

there were influencing studies in some of the estimations (sensitivity analysis). Therefore, parallel forms reliability should be addressed in greater depth in the VISA-P. Second, we found a high degree of heterogeneity in many of the analyses, although we were unable to study the causes of this through factor analysis with meta-regression.[21] Third, we have already seen that the only aspect that obtained a satisfactory quality and did not show influential studies was internal consistency. Fourth, the ICC was superior in Asians and there were no highly influential studies that altered

the results by at least 10%. Fifth, we would like to point out that the quality of the studies, except for the internal consistency, presented a significant proportion of *doubtful* and *inadequate* studies. In this sense, it is important to note that only those patient-reported outcome measures with strengthened psychometric properties are able to improve clinical research decision-making.[36] For this reason, it is important to pay attention to validation and adaptation studies, following the COSMIN recommendations to improve the clinical usefulness, because low values of reliability can lead to clinical decision errors.[38]

Finally, we were unable to study the asymmetry of funnel plots, discussing the possible causes, including publication bias.[24] In short, it seems that the VISA-P has some limitations in its reliability according to the results of our meta-analysis and with respect to its validity. Future researchers will have to assess whether the VISA-P has good quality indicators (remaining COSMIN items).[19] We must bear in mind that the questionnaire should adhere to virtually all parts of it, and studies are therefore needed to analyze the validity of the VISA-P and its cross-cultural adaptations, including the responsiveness, because of its relevant implications to clinical practice.[39]

Limitations of the Study

Although we were very rigorous in examining the articles and reviewing them in pairs and blinded, as a limitation we must mention that we found no studies analyzing the reliability of the VISA-P, either in the databases reviewed or in other sources. It is also possible that there are papers that analyze this issue, but published in other languages, which we were unable to analyze in depth and consequently were directly excluded from our review.

CONCLUSIONS

With the evidence provided in this meta-analysis using the MEDLINE, EMBASE, and Scopus databases, we can say that the reliability of the VISA-P questionnaire for assessing the severity of symptoms and the repercussions of these on daily and physical activities should be analyzed in greater depth and with more scientific evidence, because the only type of reliability that did not present problems was internal consistency. Still pending is an analysis of the causes of heterogeneity using meta-regression models and the analysis of the asymmetry of the funnel plots, which can only be performed if more papers are published analyzing the reliability of the VISA-P and always using the same comparative test for the analysis of parallel forms. Accordingly, to assess the effectiveness of an intervention, we must be cautious until new studies are conducted, because a direct correlation between the VISA-P and similar questionnaires or the variability produced by certain factors, such as the continent where the study is carried out, is not clear. Also pending is the analysis of the validity of the questionnaire and its cross-cultural adaptations, assessing the aspects contemplated in COSMIN that have not been examined in this meta-analysis (content validity, structural validity, cross-cultural validity/measurement invariance, measurement error, and criterion validity). Consequently, in view of the above arguments, to use VISA-P in routine clinical practice, new studies are needed to assess the reliability and the validity of the VISA-P, and further scientific evidence should be obtained before the systematic implementation of this questionnaire.

ACKNOWLEDGMENTS

The authors thank Maria Repice and Ian Johnstone for help with the English language version of the text.

References

1. Visentini PJ, Khan KM, Cook JL, et al. The VISA score: an index of severity of symptoms in patients with jumper's knee (patellar tendinosis). Victorian Institute of Sport Tendon Study Group. *J Sci Med Sport* 1998;1: 22–28.
2. Hernandez-Sanchez S, Hidalgo MD, Gomez A. Responsiveness of the VISA-P scale for patellar tendinopathy in athletes. *Br J Sports Med*. 2014; 48:453–457.
3. Wageck BB, de Noronha M, Lopes AD, et al. Cross-cultural adaptation and measurement properties of the Brazilian Portuguese version of the victorian Institute of sport assessment-patella (VISA-P) scale. *J Orthop Sports Phys Ther*. 2013;43:163–171.
4. Malliaras P, Cook J, Purdam C, et al. Patellar tendinopathy: clinical diagnosis, load management, and advice for challenging case presentations. *J Orthop Sports Phys Ther*. 2015;45:887–898.
5. Attia M, Scott A, Carpentier G, et al. Greater glycosaminoglycan content in human patellar tendon biopsies is associated with more pain and a lower VISA score. *Br J Sports Med*. 2014;48:469–475.
6. Herbort M, Kittl C, Mayr H. Return to play following tendon injuries. In: Musahl V, Karlsson J, Krutsch W, et al, eds. *Return to Play in Football*. Berlin, Heidelberg: Springer; 2018:661–668.
7. Medical Staff, Football Club Barcelona. Guide to clinical practice for tendinopathies: diagnosis, treatment and prevention. *Apunts Med Esport*. 2012;47:143–168.
8. Morgan S, Coetzee FF. Proposing a Patellar Tendinopathy Screening tool following a systematic review. *S Afr J Physiother*. 2018;74:454.
9. Fitzner K. Reliability and validity: a quick review. *Diabetes Educ*. 2007; 33:775–776.780.
10. Hernandez-Sanchez S, Hidalgo MD, Gomez A. Cross-cultural adaptation of VISA-P score for patellar tendinopathy in Spanish population. *J Orthop Sports Phys Ther*. 2011;41:581–591.
11. Lohrer H, Nauck T. Cross-cultural adaptation and validation of the VISA-P questionnaire for German-speaking patients with patellar tendinopathy. *J Orthop Sports Phys Ther*. 2011;41:180–190.
12. Kaux JF, Delvaux F, Oppong-Kyei J, et al. Cross-cultural adaptation and validation of the victorian Institute of sport assessment-patella questionnaire for French-speaking patients with patellar tendinopathy. *J Orthop Sports Phys Ther*. 2016;46:384–390.
13. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25: 3186–3191.
14. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterol*. 2004;126:S124–S128.
15. López-Pina JA, Sánchez-Meca J, López-López JA, et al. The yale-Brown obsessive compulsive scale: a reliability generalization meta-analysis. *Assessment*. 2015;22:619–628.
16. Green CE, Chen CE, Helms JE, et al. Recent reliability reporting practices in Psychological Assessment: recognizing the people behind the data. *Psychol Assess*. 2011;23:656–669.
17. Vacha-Haase T. Reliability generalization: exploring variance in measurement error affecting score reliability across studies. *Educ Psychol Meas*. 1998;58:6–20.
18. Moher D, Liberati A, Tetzlaff J, et al; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097.
19. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147–1157.
20. Hakstian AR, Whalen TE. A k-sample significance test for independent alpha coefficients. *Psychometrika*. 1976;41:219–231.
21. Sánchez-Meca J, López-López JA, López-Pina JA. Some recommended statistical analytic practices when reliability generalization studies are conducted. *Br J Math Stat Psychol*. 2013;66:402–425.

22. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560.
23. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011)*. The Cochrane collaboration; 2011. Available at: www.cochrane-handbook.org.
24. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. *BMJ*. 2011;343:d4002.
25. Frohm A, Saartok T, Edman G, et al. Psychometric properties of a Swedish translation of the VISA-P outcome score for patellar tendinopathy. *BMC Musculoskelet Disord*. 2004;5:49.
26. Maffulli N, Longo UG, Testa V, et al. VISA-P score for patellar tendinopathy in males: adaptation to Italian. *Disabil Rehabil*. 2008;30:1621–1624.
27. Zwerver J, Kramer T, van den Akker-Scheek I. Validity and reliability of the dutch translation of the VISA-P questionnaire for patellar tendinopathy. *BMC Musculoskelet Disord*. 2009;10:102.
28. Park BH, Seo JH, Ko MH, et al. Reliability and validity of the Korean version VISA-P questionnaire for patellar tendinopathy in adolescent elite volleyball athletes. *Ann Rehabil Med*. 2013;37:698–705.
29. Korakakis V, Patsiaouras A, Malliaropoulos N. Cross-cultural adaptation of the VISA-P questionnaire for Greek-speaking patients with patellar tendinopathy. *Br J Sports Med*. 2014;48:1647–1652.
30. Çelebi MM, Köse SK, Akkaya Z, et al. Cross-cultural adaptation of VISA-P score for patellar tendinopathy in Turkish population. *Springerplus*. 2016;5:1453.
31. Hernandez-Sanchez S, Abat F, Hidalgo MD, et al. Confirmatory factor analysis of VISA-P scale and measurement invariance across sexes in athletes with patellar tendinopathy. *J Sport Health Sci*. 2017;6:365–371.
32. Acharya GU, Kumar A, Rajasekar S, et al. Reliability and validity of Kannada version of Victorian Institute of Sports Assessment for patellar tendinopathy (VISA-P-K) questionnaire. *J Clin Orthop Trauma* 2019;10(suppl 1):S189–S192.
33. Çelik D, Arıkan F, Gözbek A. Translation and cross-cultural adaptation and version of the victorian Institute of sport assessment-patella (VISA-P) scale to Turkish. *Orthop J Sports Med*. 2018. doi: 10.1177/2325967117S00100. Available at: https://journals.sagepub.com/doi/pdf/10.1177/2325967117S00100.
34. Blazina ME, Kerlan RK, Jobe FW, et al. Jumper's knee. *Orthop Clin North Am*. 1973;4:665–678.
35. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–483.
36. Frost MH, Reeve BB, Liepa AM, et al; Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group; What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007;10(suppl 2):S94–S105.
37. Cohen J *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Psychology Press (Taylor & Francis Group); 1988:413.
38. Charter RA. A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *J Gen Psychol*. 2003;130:290–304.
39. Kyte DG, Calvert M, van der Wees PJ, et al. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*. 2015;101:119–125.