Validation of Patient-Reported Outcome Measures in Orthopedics

Validación de medidas de resultados informados por los pacientes en Ortopedia y Traumatología

Catalina Vidal[1]   María Jesús Lira[1]   Pablo Besa[1]   Maximiliano Carmona[1]   Sebastián Irarrázaval[1]

[1] Department of Orthopedics and Traumatology, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Región Metropolitana, Chile

Rev Chil Ortop Traumatol 2022;63(1):e55–e62.

Address for correspondence Dr. Sebastián Irarrázaval Domínguez, MD, Departamento de Ortopedia y Traumatología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Diagonal Paraguay 362, Santiago, Región Metropolitana, 8330077, Chile (e-mail: sirarraz@med.puc.cl).

Abstract

Keywords
► clinical outcomes
► patient-reported outcomes
► PROMs
► methodology
► validation

In recent years, there has been an increase in the use of questionnaires designed to measure outcomes in the medical practice. To use a questionnaire in a population different from the one for which it was originally created and designed, it is necessary to carry out a rigorous adaptation process, with a certain methodology. The objective of the present methodological guide is to describe the process of translation, cross-cultural adaptation, and validation of patient-reported outcome measures in Orthopedics and Traumatology.
Level of evidence: IV.

Resumen

Palabras clave
► resultados clínicos
► resultados informados por los pacientes
► MRIPs
► metodología
► validación

En los últimos años, ha habido un aumento en la aplicación de cuestionarios diseñados para la medición de resultados (o desenlaces) clínicos en la práctica médica. Para aplicar un cuestionario en una población distinta a la cual fue originalmente creado y diseñado, es necesario llevar a cabo un proceso riguroso de adaptación, con una determinada metodología. El objetivo de esta guía metodológica es describir el proceso de traducción, adaptación transcultural y validación de medidas de resultados informados por los pacientes (MRIPs) en Ortopedia y Traumatología.
Nivel de evidencia: IV

Introduction

In recent years, as health care has migrated to patient-centered care, there has been a drastic increase in the use of questionnaires designed to measure clinical outcomes reported by the patient.[1,2] These tools mainly enable the determination of the changes associated with therapeutic interventions, as well as the follow-up and prognosis of different pathologies, with the goal of performing the evaluation through a score. The results obtained through these

questionnaires are also known as Patient-Reported Outcome Measures (PROMs).[3]

Specifically in the field of Orthopedics and Traumatology, PROMs have been fundamental to evaluate health aspects that are relevant to patients, their families and their support network.[3] Today, they are widely used in the clinical practice, especially in patients undergoing surgical interventions, enabling the collection of information on the impact on the functionality of activities of daily living, self-care, symptoms, and quality of life.[4] Moreover, they have an impact on scientific research, since they enable the comparison of treatments in a standardized way.[1]

These questionnaires can accurately measure the outcome of interest, provided they are used by the populations for which they were designed and tested.[5] In the event that these questionnaires are to be used in other populations, it is necessary to previously carry out a process of translation, cultural adaptation, and validation, which must be performed following a specific methodology. This workflow seeks to guarantee that the adapted questionnaire has the same characteristics as the original, in such a way that it is capable of correctly interpreting the results obtained in each use, avoiding errors of classification, diagnosis, or decision-making.[6]

In this context, the objective of the present methodological guide is to describe the process of translation, cross-cultural adaptation, and validation of questionnaires for clinical results in Orthopedics and Traumatology.

Translation, cross-cultural adaptation and validation of questionnaires

The translation, cross-cultural adaptation and validation of questionnaires should be carried out in those cases in which one wants to use a questionnaire that measures clinical results in a population with a language and/or culture different from those for which it was designed.

The first action is to contact the authors who originally designed the questionnaire and request permission to start the translation and validation process. The steps that must be followed, once the process is authorized, are described as follows:[4,7]

I. Translation and cross-cultural adaptation; and
II. Evaluation of psychometric properties.

I. Translation and cross-cultural adaptation

It is important to bear in mind that a process of translation and cross-cultural adaptation will be necessary whenever the questionnaire is intended to be used in a country other than the country of origin, even in cases in which the language is the same or similar.[8]

The translation process is divided into four steps: *1. translation; 2. synthesis; 3. back translation; 4. committee of experts* (►**Table 1**).

The initial translation must be performed based on the questionnaire in its original language by at least two independent translators who must be native in the original language. Then, a consensual version, which is called a synthesis, is generated based on the two translations. Next, in the back-translation step, the synthesis is translated back to the original language of the questionnaire, to check that there are no relevant discrepancies regarding the original tool. After this review, the synthesis version is reviewed by an expert committee made up of the translators, a methodologist, a linguist, the research team, and other people who can contribute to the review, such as community representatives (►**Figure 1**). And so, we reach what we will call the preliminary version of the questionnaire.

The preliminary version should be evaluated by 30 to 40 patients.[4] Although some studies show that this process can be carried out in any population,[8] other authors[4,7] point out that it is recommended that it should be carried out by patients who have the pathology or painful syndrome evaluated by the questionnaire, since this enables a better approach and generates better feedback for the next steps.

Table 1 Four steps in the the translation of a questionnaire[4,7]

| Step	Name	Task	Participants
1	*Translation*	Two translators, whose mother tongue is the original language of the questionnaire, must perform the translation from the original language to the target language separately. Thus, two different translations are are obtained	Translator 1 Translator 2
2	*Synthesis*	A consensus is reached between the two translations carried out, resolving the discrepancies	Translator 1 Translator 2
3	*Back translation*	Two translators, whose mother tongue is the target language of the questionnaire, must perform the translation to the original language based on the consensus version separately. Thus, two different translations are obtained	Translator 1 Translator 2
4	*Committee of experts*	The translations are reviewed and it is observed that the back translation is similar to the original questionnaire. Discrepancies from the consensus version are reviewed, and a preliminary version is produced	Translators[2–4] Methodologist Linguist Research team Experts in the field Others
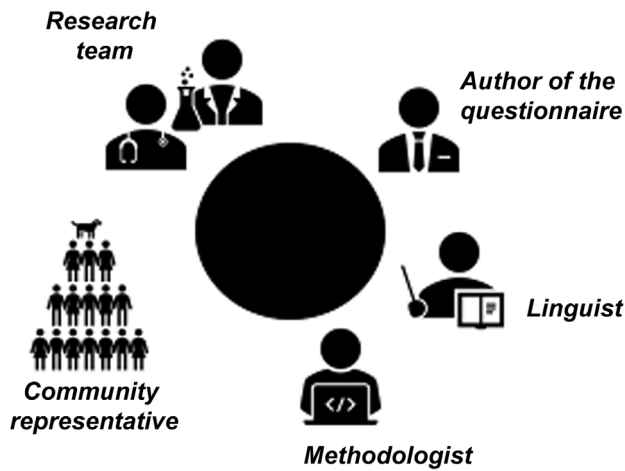
Fig. 1 Members of the expert committee for the pre-final version of the questionnaire.

Special attention must be paid when obtaining a representative population in terms of their level of schooling. According to the 2016-2017 National Health Survey held by the Chilean Ministry of Health, 77% of the Chilean population has less than 12 years of schooling, and 23.7%, less than 8 years, so the target population should be a representative sample of that population.[9]

The questionnaires must be evaluated respecting the properties with which they were created. In the case of most questionnaires for clinical results in Orthopedics and Traumatology that collect PROMs, they are designed to be self-administered by patients, so, at this stage of the evaluation, it is important that they fill out the questionnaire themselves, with instructions to answer all of the questions. On the other hand, if the questionnaire has been designed to be administered by a third party, it is most appropriate that all the questionnaires be administered by the same person.[10]

After the application of the preliminary questionnaire, a semi-structured interview is carried out, in which the patient is invited to comment globally on what he or she thought of the questionnaire, and is asked about difficulties or conflicts when trying to answer any item in detail. The meaning attributed by the patient to each question should be evaluated to ensure that the final version maintains its equivalence to the original version. This interview can be recorded, so all responses can be tabulated later in order to identify those items that generated conflict in more than 15% of the respondents.[6]

Finally, a new meeting of the expert committee is planned with the aim of discussing these questions and making the necessary modifications to develop the final version (►**Figure 2**).

Undoubtedly, this process, which requires great effort, is necessary to be able to produce a version that is similar to the original version. Following these guidelines brings us closer to having a tool that can be used reliably in a population. However, despite the fact that up to this point we have obtained very useful information when comprehending the process of understanding the questionnaire, this does not provide information on validity or reliability.[11,12]

II. Evaluation of psychometric properties

The following steps must also comply with a rigorous process, in which the psychometric properties of the questionnaire will be sought, which will consist in the evaluation of validity and reliability.[11,13,14] In this process, the questionnaire in its final version must be used by a representative sample of the population at whom it is aimed (►**Table 2**).[15–23]

Validity

The validity of a questionnaire corresponds to the ability to adequately measure what it wants to measure, and correctly evaluate the characteristic for which it was created.[6] Three types of validity are defined:
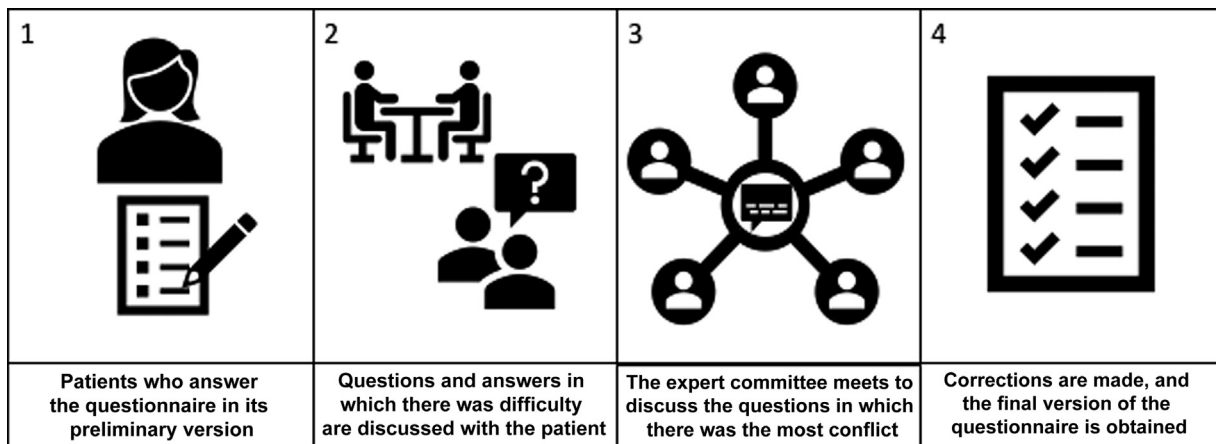

Fig. 2 Process for obtaining the final version of the questionnaire.

Table 2 Examples in Orthopedics and Traumatology

Item	Description		
Subjects	Representative sample of the population in whom I intend to use this tool. Sample size of 5 to 10 subjects for each question contained in the questionnaire[15,16]		
	Example 1	*The Arthritis Research UK Musculokeletal Health Questionnaire (MSK-HQ) is a recently developed PROM that assesses quality of life related to musculoskeletal health. In its creation, it was validated with a sample of 570 patients with osteoarthritis. In 2019, in the same country, a study was conducted to evaluate the validity and reliability of the MSK-HQ in people with inflammatory arthritis.[17]*	
	Example 2	*The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) is a PROM that is widely used to assess pain, stiffness, and function in patients with osteoarthritis of the hip and knee.[18] Although the WOMAC has also been used to assess postoperative foot and ankle surgery, it was not validated for this purpose. Ponkilainen et al. carried out the validation process in 2019.[19]*	
Instruments and benchmark tests	Use the tool to be validated together with the benchmark tests, which may be a similar questionnaire, a clinical examination, a laboratory examination, imaging etc.		
	Example 1	*In a questionnaire created and validated by Schnetzke in 2016, the objective was to evaluate the range of motion of the patient's wrist and elbow. The self-administered tool contained questions about the ability to perform certain movements, and the goniometric evaluation of a trained examiner was used as a reference test for the validation process.[20]*	
	Example 2	*Pardis et al. performed the validation of the Athlete Disability Index (ADI), a questionnaire to assess disability associated with low back pain in athletes. For the process, they used this questionnaire in conjunction with the Oswestry Disability Index (ODI), and the Roland-Morris Disability Questionnaire (RDQ), both tools originally used to evaluate low back pain.[21]*	
Reevaluation	The application of the evaluations must be repeated in a certain time		
	Example 1	*In Spain, the SEROD group carried out the validation of the Hip and knee questionnaire, which assesses the impact of musculoskeletal pathologies on the quality of life of patients. To do this, they included patients who underwent total knee arthroplasty and used the questionnaire together with the benchmark test, before surgery and 6 months after surgery.[22]*	
	Example 2	*In Korea, the Core Outcome Measures Index questionnaire was validated in patients with degenerative lumbar pathology. The questionnaire was used by a group of patients in their first consultation, in conjunction with the reference test, and after 2 weeks it was used again in conjunction with a transition question (no changes, slight changes, moderate changes, many changes). This correlated with changes in the questionnaire.[23]*	

1. *Content validity*: it refers to the extent to which the items in the questionnaire are representative of the characteristic that you want to measure or for which it is designed. This process is crucial in the stage of development of the questionnaire; however, it is also crucial at the time of validation. A panel of experts in the clinical outcome to be measured has the task of evaluating the validity of the content.[12,24,25]

2. *Construct validity*: it assesses the degree to which the questionnaire reflects the characteristic or concept that it wants to measure,[26] estimating its association with other variables (or measures of a construct) with which it should have a positive, negative or null correlation. It is worth mentioning that the construct is defined as the "*underlying theory in the phenomenon or concept to be measured. It is an unobservable quality in a population of subjects*".[6]

3. *Criterion validity*: it corresponds to the relationship of the score of each subject with a gold standard that measures the same characteristic.[6,8] This validity is made up of two dimensions:

   - Concurrent or convergent validity: the degree to which the result of the questionnaire agrees with some standard at a given moment in time; and
   - Predictive validity: the degree to which it is able to predict a certain result.

Reliability

The reliability of a questionnaire is the consistency of its results, which can be evaluated using internal consistency, intraobserver reliability, and interobserver reliability[12]:

1. *Internal consistency*: it reflects the degree to which the items in the questionnaire are correlated, or if they are consistent in measuring the same phenomenon. Internal consistency is commonly estimated using the alpha coefficient, also known as Cronbach alpha.[27] The Cronbach alpha ranges from 0 to 1, and a number close to 0 indicates that there is no internal consistency, or that the items are not correlated. A number close to 1 indicates perfect internal consistency. A cut-off point for adequate consistency has been stated to be 0.7.[28] (►**Appendix 1**)
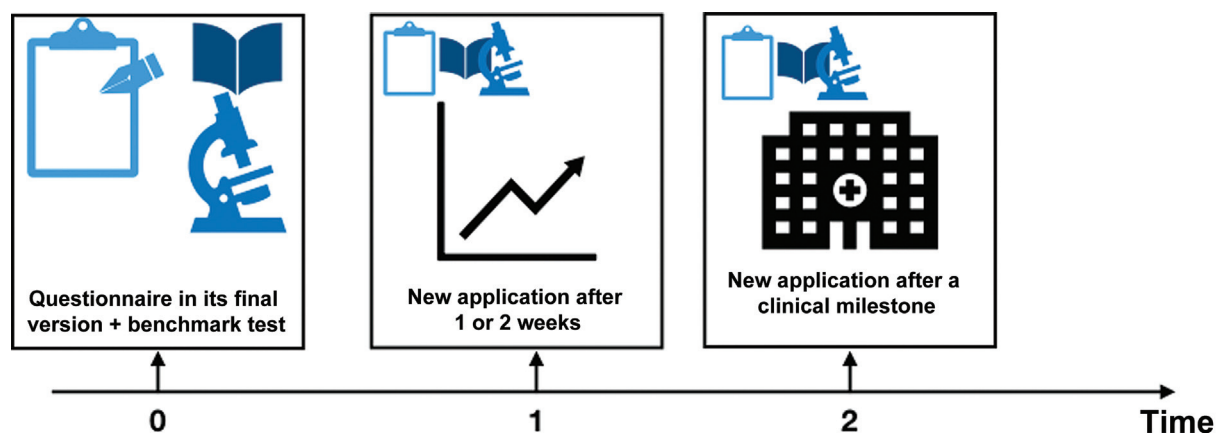
Fig. 3 Steps to obtain the psychometric properties of the questionnaire.

2. *Intraobserver reliability*: it corresponds to a way of measuring the stability of the questionnaire scores, in the same subjects and with the same method, at different times.[6,12,14] This will enable the evaluattion of the level of agreement in the responses of the subject at two or more intervals of time (►**Figure 3**). This process can be carried out in two ways:
   - With an interval of one to three weeks from the first application, expecting that there is no significant change in the responses, because no significant changes in the patient's condition are expected either; and
   - According to a clinical milestone in which a result is expected. For example, using the questionnaire before and after a surgical intervention through which improvement is expected.

3. *Interobserver reliability*: this property assesses the concordance or agreement between 2 evaluators who apply the same tool to the same subject,[6] in the event that the questionnaire is not self-administered.

It is important to mention that not all the items of validity and reliability correspond to the process of all the questionnaires in Orthopedics and Traumatology. It is paramount to evaluate, in each case, the steps that are required for the correct development of this stage.

Conclusions

In recent years, PROMs have made it possible to improve the quality of patient care,[29,30] especially in the field of Orthopedics and Traumatology, due to the special attention dedicated to results related to pain, functionality and quality of life.[10]

Translation, cross-cultural adaptation, and validation can be long, difficult or expensive processes; however, they are essential when using a PROM in the clinical practice. This process is necessary even when one wishes to apply a questionnaire in countries with the same language.[6] Sometimes it is assumed that cultural adaptation to a different language guarantees the psychometric properties of the questionnaire, which can lead to errors in evaluations. It is necessary to complete the validation process rigorously, measuring the validity and reliability of the measurement tool.

Given the importance of the use of PROMs adapted and validated in the clinical practice to improve the control and follow-up of patients, the Organisation for Economic Co-operation and Development (OECD) has launched an initiative for the systematic collection of PROMs.[31] In the field of Orthopedics and Traumatology, this initiative promotes the evaluation of patients undergoing elective hip and knee arthroplasty prior to surgery, and 6 and 12 months after surgery.[32] The creation of a network for the systematic collection of PROMs in all countries will help to investigate the determinants of quality in medical care, to make national and international comparisons, and to align the practice with health policies.[31]

Therefore, it is necessary to promote the use of PROMs in the different health problems associated with the specialty of Orthopedics and Traumatology, using tools that have been translated, adapted and validated to the corresponding population, in order to make their use part of the routine of the clinical practice and thus access a common and objective language. (►**Appendix 2**)

Conflict of Interests
Dr. Sebastián Irarrázaval reports that he is an Associate Editor of *Revista Chilena de Ortopedia y Traumatología*. None of the other authors have any conflict of interests to declare.

References

1 Evans JP, Smith A, Gibbons C, Alonso J, Valderas JM. The National Institutes of Health patient-reported outcomes measurement information system (PROMIS): a view from the UK. Patient Relat Outcome Meas 2018;9:345–352
2 World Health Organization, Organisation for Economic Co-operation and Development, The World Bank. Delivering quality health services: a global imperative for universal health coverage.

Geneve: WHO; 2018. Available from: https://www.worldbank.org/en/topic/universalhealthcoverage/publication/delivering-quality-health-services-a-global-imperative-foruniversal-health-coverage

3 Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. Qual Life Res 2008;17(09):1125–1135

4 Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the cross-cultural adaptation of the DASH & QuickDASH outcome measures. Institute for Work & Health. 2007;1(01):1–45

5 Ruzbarsky JJ, Marom N, Marx RG. Measuring quality and outcomes in sports medicine. Clin Sports Med 2018;37(03):463–482

6 Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. Salud Publica Mex 2013;55(01):57–66

7 Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the cross-cultural adaptation of health status measures. New York. J Am Acad Orthop Surg 2002;12:1–9

8 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46(12):1417–1432

9 MINSAL. Primeros y segundos resultados de ENS 2016-2017. 2018 [Revisado el 1 julio 2020] en: http://epi.minsal.cl/resultados-encuestas/

10 Gagnier JJ. Patient reported outcomes in orthopaedics. J Orthop Res 2017;35(10):2098–2108

11 Lauffer A, Solé L, Bernstein S, Lopes MH, Francisconi CF. Cómo minimizar errores al realizar la adaptación transcultural y la validación de los cuestionarios sobre calidad de vida: aspectos prácticos. Rev Gastroenterol Mex 2013;78(03):159–176

12 Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth 2017;11(Suppl 1):S80–S89

13 Pasquali L. Instrumentos psicológicos: manual prático de elaboração. Brasília: LabPAM/IBAPP; 1999

14 Pasquali L. Psychometrics. Rev Esc Enferm USP 2009;43:992–929

15 Gorusch RL. Factor Analysis. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1983

16 Pedhazur RJ. Multiple Regression in Behavioral Research: Explanation and Prediction. Fort Worth, TX: Harcourt Brace College Publishers; 1997

17 Norton S, Ellis B, Santana Suárez B, et al. Validation of the Musculoskeletal Health Questionnaire in inflammatory arthritis: a psychometric evaluation. Rheumatology (Oxford) 2019;58(01):45–51

18 Ackerman I. Western ontario and mcMaster universities osteoarthritis index (WOMAC). Aust J Physiother 2009;55(03):213

19 Ponkilainen VT, Häkkinen AH, Uimonen MM, Tukiainen E, Sandelin H, Repo JP. Validation of the Western Ontario and McMaster Universities Osteoarthritis Index in Patients Having Undergone Ankle Fracture Surgery. J Foot Ankle Surg 2019;58(06):1100–1107

20 Schnetzke M, Schüler S, Keil H, et al. Development and validation of a novel questionnaire for self-determination of the range of motion of wrist and elbow. BMC Musculoskelet Disord 2016;17(01):312

21 Noormohammadpour P, Hosseini Khezri A, Farahbakhsh F, Mansournia MA, Smuck M, Kordi R. Reliability and validity of athletes disability index questionnaire. Clin J Sport Med 2018;28(02):159–167

22 Castellet E, Ares O, Celaya F, et al; SEROD group. Transcultural adaptation and validation of the "Hip and Knee" questionnaire into Spanish. Health Qual Life Outcomes 2014;12(01):76

23 Kim HJ, Yeom JS, Nam Y, et al. Validation and cross-cultural adaptation of the Korean version of the Core Outcome Measures Index in patients with degenerative lumbar disease. Eur Spine J 2018;27(11):2804–2813

24 Lawshe CH. A quantitative approach to content validity. Person Psychol 1975;28:563–575

25 Barrett RS. Content validation form. Public Pers Manage 1992;21(01):41–52

26 Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52(04):281–302

27 Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334

28 Nunnally J. Psychometric Theory. New York: McGraw-Hill; 1978

29 Prodinger B, Taylor P. Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples. BMC Health Serv Res 2018;18(01):87

30 Peters RM, van Beers LWAH, van Steenbergen LN, et al. Similar superior patient-reported outcome measures for anterior and posterolateral approaches after Total hip arthroplasty. J Arthroplasty 2018;33(06):1786–1793

31 Organisation for Economic Co-operation and Development (OECD) PatientReported Indicators Survey (PaRIS). 2017. Available from: http://www.oecd.org/health/paris.htm

32 Canadian Institute for Health Information, Organisation for Economic Co-operation and Development. OECD Patient-Reported Indicator Surveys (PaRIS) Initiative: Patient-Reported Outcome Measures (PROMs) for Hip and Knee Replacement Surgery — International Data Collection Guidelines. Ottawa, ON: CIHI; 2019

33 Guic E, Galdames S, Rebolledo P. Adaptación cultural y validación de la versión chilena del Cuestionario de Discapacidad Roland-Morris. [Validation and cultural adaptation of the Chilean version of the Roland-Morris Disability Questionnaire]Rev Med Chil 2014;142(06):716–722

34 Keller A, Wagner P, Izquierdo G, et al. Cross-cultural adaptation and validation of the VISA-A questionnaire for Chilean Spanish-speaking patients. J Orthop Surg Res 2018;13(01):177

35 Pellegrini MJ, Poniachik R, Nuñez A, Escudero MI, Carcuro G, Cortes AA. Cross-cultural adaptation and validation of the Foot and Ankle Outcome Score (FAOS) into Spanish (Chile). Foot Ankle Surg 2020;26(07):790–796

36 Vera-Villarroel P, Silva J, Celis-Atenas K, Pavez P. Evaluación del cuestionario SF-12: verificación de la utilidad de la escala salud mental. Rev Med Chil 2014;142(10):1275–1283

Appendix 1 Cronbach alpha

The Cronbach alpha is calculated by correlating the score of each questionnaire item with the total score of each observation (respondents), and then comparing it with the variance of all the scores of the individual items:

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_x^2}\right)$$

$k =$ number of items in the questionnaire
$\sigma_x^2 =$ variance associated with item i
$\Sigma\sigma_x^2 =$ sum of variances of each item
$\sigma_{y1}^2 =$ variance associated with the total observed score (sum of items)

This calculation can be performed using software such as Excel or statistical analysis software such as R, Stata or SPSS, among others.

An example is given below in which the data of 3 subjects are obtained (ID01, ID02, ID03) in a 2-item questionnaire.

Subject	Item I	Item II	Sum
ID01	5	4	9
ID02	7	8	15
ID03	6	7	13
Variance	1	4.3	9.3

$k = 2$
$\Sigma\sigma_x^2 = 5.3$
$\sigma_{y1}^2 = 9.3$

$$\sigma_{y1}^2 = 9.3$$
$$\alpha = \left(\frac{2}{1}\right)\left(1 - \frac{5.3}{9.3}\right) = 0.43$$

A value of 0.43 indicates a Cronbach alpha below the appropriate value. Therefore, in this example, low internal consistency is observed for the questionnaire.

Appendix 2: Questionnaires used in Orthopedics and Traumatology and validated in Chile

Questionnaire name	Title of the published article	Journal	Year
Roland-Morris	Validation and cultural adaptation of the Chilean version of the Roland-Morris Disability Questionnaire[33]	*Curr Pharm Teach Learn*	2018
VISA-A questionnaire	Cross-cultural adaptation and validation of the VISA-A questionnaire for Chilean Spanish-speaking patients[34]	*J Orthop Surg Res*	2018
FAOS	Cross-cultural adaptation and validation of the Foot and Ankle Outcome Score (FAOS) into Spanish (Chile)[35]	*Foot and Ankle Surgery*	2020
SF-12	Evaluación del cuestionario SF-12: verificación de la utilidad de la escala salud mental[36]	*Rev Med Chile*	2014

Note: The search was carried out in the Pubmed, Scielo and Google Scholar databases, looking for original research articles published until October 2020. Keywords: validation, questionnaire, PROMs, orthopedics, Chile.