Research article | Open Access

# Prospective, Multicenter Study Protocol to Adaptate and Validate the Simple Shoulder Test to Spanish Population

**Medina-porqueres I**[1*], **Cantero-tellez R**[1], **Gomez-calvente M**[1] and **Matsen Fa**[2]

[1]University of Malaga, Faculty of Health Sciences, Department of Physical Therapy, Malaga, Spain

[2]University of Washington, Department of Orthopaedics, Seattle, WA, USA

**\*Corresponding author**: Medina-Porqueres I, University of Malaga, Faculty of Health Sciences, Department of Physical Therapy, Malaga, Spain, Tel: 951 952 858; E-mail: imp@uma.es

## Abstract

**Background:** The validation of widely used scales facilitates the comparison across international patient samples. The Simple Shoulder Test (SST) was developed for measuring functional limitations of the affected shoulder. The SST has been adopted both for clinical practice and research purposes, although cross-cultural differences may impose difficulties in its use.

**Objective:** The aim of this study is to translate and cross-culturally adapt the SST into Spanish, to obtain a reliable and feasible SST-Spanish version, and to evaluate the psychometric properties of this tool in multiple trauma centres.

**Design:** Cross-cultural and validation study protocol.

**Setting:** Five public and private hospitals in Spain (Costa del Sol Marbella Hospital, Quiron Malaga Hospital, Malaga TECAN Center, Virgen de la Victoria Malaga Hospital, Lasalle Functional Rehabilitation Institute from Madrid, Santa Coloma de Gramanet Hospital, Valencia Moviment I Salut Center, and Manacor Hospital).

**Method:** Patients of any gender attending will be recruited. The procedures of translation and cross-cultural adaptation to Spanish will be conducted following proposed guidelines that involved translation, synthesis of translations, back translation, committee review and testing of the pre-final version.

**Conclusion:** The Spanish version of the SST questionnaire will be obtained, and its comprehensibility and usefulness in shoulder patients in both public and private settings will be evaluated.

**Keywords:** Simple shoulder test; Cross-cultural adaptation; Validation; Spanish; Psychometric properties; Clinimetry

## Introduction

Shoulder pain is the second more common musculoskeletal disorder in the primary care setting. Actually, it is estimated that 20% of people will suffer from shoulder pain at some point in their lives [1]. More than 6 million people visit the orthopaedic specialist a year in the U.S. [2]. Shoulder disorders may cause pain and/or reduced joint mobility, which necessarily affect quality of life of individuals; and, its influence on productivity and total number of worked hours has a major, well known socio-economic impact [3].

Conditions affecting the shoulder should be evaluated based on how activities of daily living might be affected. One of the numerous evaluating instruments is the Simple Shoulder Test (SST), which has not yet been translated and validated to Spanish population. The evidence suggests that SST is one of the first options for measuring disability and function in patients with shoulder problems, showing good results in terms of validity, reliability and responsiveness [4,5].

The Simple Shoulder Test (SST) is a tool designed to evaluate functional limitations of an injured shoulder that compromise an individual's daily activities [6]. It is a shoulder function scale consisting of 12 items that ask people about their ability to tolerate or perform 12 activities of daily living (ADL). The individual indicates that he or she is able or not to do the activity. The scores range from 0 (worst) to 100 (best) and are reported as the percentage of answered items to which the person responds in the affirmative. It is a practical method for assessment of shoulder function before and after treatment.

SST is widely used since the ease of its application has facilitated the comparison of patient outcomes. It is a standardized instrument developed to systematically document shoulder function. SST is also a helpful indicator of the time required to reach a maximum benefit of a treatment for shoulder pain [7]. The questionnaire was developed based on common patient complaints presented to practitioners. Numerous interventional studies have used this scale for the assessment of shoulder function [8-10], which reinforce the need for cultural translation and validation to other languages, facilitating assessment of shoulder conditions among different populations groups.

This study is designed to describe the process to translate and culturally adapt the SST into Spanish, to obtain a reliable and feasible Spanish version of SST, and to evaluate the psychometric properties of

this tool in multiple trauma centres. External validation, through a multi-centre study, will be needed to ensure the reliability, content validity and construct validity of the measure hold across different settings and different participants. This would increase the generalizability and comparability of the results and thereby increase the value of the questionnaire as a quality improvement tool.

## Methods

### Setting and context

Patients of any gender attending any of the following hospitals (Costa del Sol Marbella Hospital, Quiron Malaga Hospital, Malaga TECAN Center, Virgen de la Victoria Malaga Hospital, Lasalle Functional Rehabilitation Institute from Madrid, Santa Coloma de Gramanet Hospital, Valencia Moviment I Salut Center, and Manacor Hospital) will be recruited and all of them will be informed about the procedure.

### Participants

Data about age, gender, level of education, employment and chronification of the illness will be collected. Table 1 summarizes inclusion and exclusion criteria.

| Inclusion criteria | | Exclusion criteria | |
|---|---|---|---|
| - | Adult patients (>18) | - | Surgery |
| - | Spanish | - | Fractures |
| - | Shoulder pain | - | Frozen shoulder |
| - | Visiting the orthopaedic clinic | - | Problems with |
| - | Written informed consent | | reading/understanding Spanish |
| | | | language |

**Table 1:** Inclusion and exclusion criteria/ Eligibility criteria.

### Ethics

The SST authors6 were contacted, and they provided authorization to conduct this study. Approval was obtained from the Research Ethics Committee of Costa del Sol Hospital prior to the initiation of this project (CODE: 011_ marzo_PR – Protocolo Validación Hombro). All study participants will provide written informed consent prior to enrolment in the study. The study will be conducted in accordance with medical professional codex and the Helsinki Declaration.

### Design

We will apply the recommended methodology for the translation and cultural adaptation of Health Related Quality of Life (HRQoL) questionnaires [11] used in others studies including direct and inverse translation and cognitive interviews [12]. An overview of the translation used and cultural adaptation processes are described in Figure 1.

### Phase 1: Forward translation

Three bilingual translators whose native language is Spanish will translate the SST from English into Spanish. Two of the translators will

be aware of the concepts on the questionnaire. The third translator will be neither aware of nor informed about the conceptual content. All translators will have expertise in cross-cultural translation scale study design and are fluent in both Spanish and English. Each translator independently will produce a forward translation of the original items, instructions and response options [11]. All the measures will be transformed into the metric system (gallons and pints to meters and pounds to kilograms).
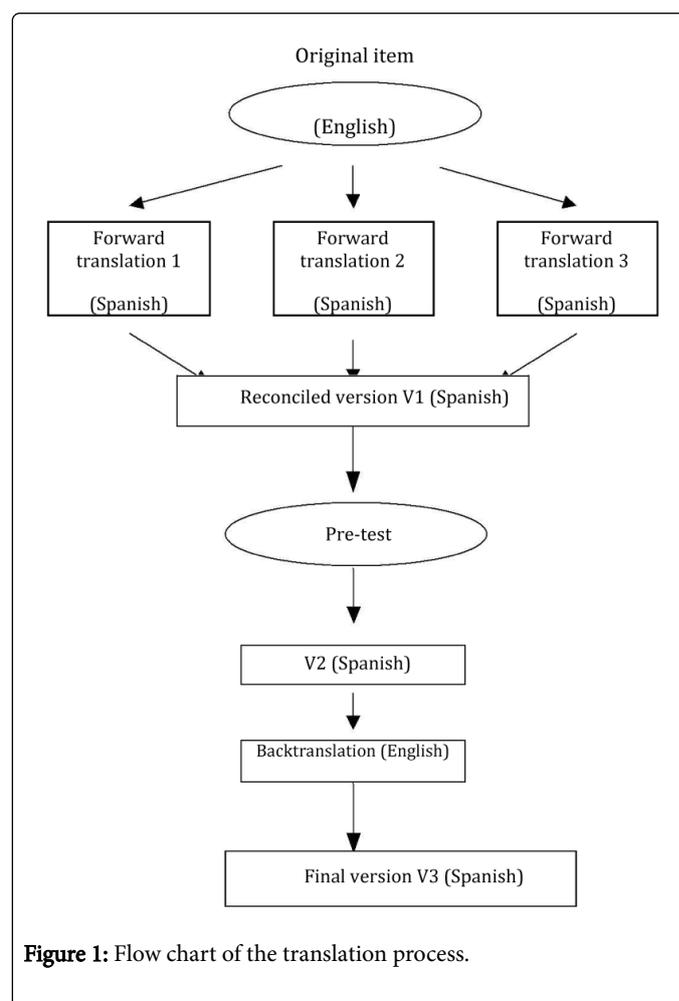


**Figure 1:** Flow chart of the translation process.

To produce a combined version (version 1) both translators and one local project manager will discuss the three translations and will agree on a single version with the aim to produce a conceptually, semantic and easy to understand equivalent translation of the original questionnaire [13]. This process will lead to additional changes to the original version where words or concepts could be untranslatable, or where words or terms could have a specific meaning in one language but a semantically different or secondary meaning in the Spanish language.

### Translation synthesis

This stage will consist of the synthesis of all three translations. This synthesis process will be fully documented. All disagreements will be resolved through discussion, ultimately reaching a consensus.

## Phase 2: Patient testing using cognitive interviews

The pre-final version of the Spanish SST will be administered in a group of 30 patients (patient testing) who have an appointment at the orthopaedic clinic because of shoulder pathologies. This number of patients has been considered a sufficient number for the pre-test showing good results [11,14]. This phase is aimed at certifying whether the patients understood the meaning of the questions presented in the SST questionnaire, whether the translation (items, instructions and responses options) is acceptable and easy to understand. This will be tested by means of cognitive interviews using "probing and paraphrasing" methodology to provide patient feedback in respect to errors or misunderstandings produced by the translation process [15]. Such cognitive interview techniques are known to minimise measurement error introduced by the translation process and enable respondent misunderstandings to be rectified [16].

Cognitive interviews will be face to face and will be conducted in an egalitarian manner by a native Spanish speaker with 30 adults aged over 18 years old, and findings will be collated and stratified using gender (male or female) and ailment (healthy or shoulder pain).

The interviews will consist of:

1)    An evaluation of the ease of comprehension of each item using dichotomous response options of either: 1) clear and comprehensible or 2) difficult to understand.

2)    An evaluation of the ease of comprehension of each item using a numerical rating scale from 0 to 10 (0 very easy to understand to 10 very difficult to understand).

3)  An investigation of individuals' interpretations of SST items with suggestions for improvements by asking those interviewed patients to express in their own words the perceived meaning of each item and then to re-phrase each item to verify their understanding.

Where problems are identified, alternative linguistic changes will be proposed and following this process version 2 of the questionnaire will be obtained. Uncertainties reported by 15 or 20% or more of the sample will provoke the revision of the questionnaire [17]. Whether the percentage is greater than 15% as if more subjects are included, the translated and adapted version of the questionnaire or instrument will be changed and a new pre-test will be conducted.

## Phase 3: Back-translation

The final phase will be the back-translation of the Spanish version 2 of the SST into English using a local professional translator, who will be a native speaker of English and fluent in Spanish, and will be blinded to the original English version of the SST questionnaire. The back-translated SST will be then compared to the original by the local project manager and the author of the original English SST to detect any misunderstandings or inaccuracies in the translation process [11].

The used translation methodology will be designed in order to reduce the cultural and social bias that may have resulted if only one translator was responsible for the translation, and aimed to ensure that the final version obtained had conceptual and semantic equivalence to the English SST with respect to the items, instructions and response options, as Beaton et al. stated [14].

## Expert committee

The committee will be constituted by methodologists, health professionals, translators involved during the process, language professionals and original author of the questionnaire [14]. The translation synthesis and back-translation versions of the SST will be submitted to the expert committee, which will review all translations and will attempt to reach a consensus regarding differences identified in the process. The main guiding principle will be that the final test should make it easy for an ordinary individual (12 years old) to understand it [14,18].

## Phase 4: Validation of the questionnaire

International Quality of Life Assessment (IQOLA) as well as other significant number of investigations supports that the study of psychometric properties is a key point in the questionnaire validation [19]. Current evidence supports a sample size of 200-300 patients to be enough in order to obtain good results in psychometric properties during validation process [20, 21]. For the purpose of this study, the sample size will be calculated based on a minimum subject/item ratio×12=12 of 10:1 (10 0) [22]. Assuming an estimated drop-out rate of 20%, 145 individuals will be recruited.

The study of psychometric properties is based on reliability and validity evaluation.

Reliability is the degree in which the obtained results are the ones we are waiting for, precise and without errors. It depends mainly on stability and internal consistency [23].

Intra-rater or test-retest reliability will be used to measure stability. Test-retest reliability consists of a repeated measurement of the same population by the same rater but at different times. To test reproducibility or test-retest reliability, patients will be asked to answer the questionnaire again within 2-7 days to see whether they completed it with the same answers. The reproducibility will be investigated by calculating the intra-class correlation coefficient (ICC), in a 2-way random model for agreement, between the test and re-test [24].

Internal consistency shows the coherence between the items of the questionnaire25. Common points the SST measures are the suffered limitations by patients with shoulder problems. This evaluation will be done using Cronbach's alpha. Values higher than 0.70 are considered indicators of good internal consistency [18,25], using a sample big enough of 30-40 patients to prove it [26].

Validity is an index of how well a test measures what it is supposed to measure. It reflects the degree in which a questionnaire measures what it is supposed to measure, that is to say, the measurement in which an instrument will take us into valid conclusions [23]. Several dimensions of validity will be studied.

## Construct validity

When a questionnaire is applied to almost two situations or populations in different scenarios, results should show differences; these existing differences are evaluated by construct validity [18]. Responsiveness is a special type of construct validity which measures the degree in which a questionnaire is sensitive enough to detect the least clinically relevant difference [18].

In the SST two groups of patients with problems in the shoulder at different levels of the ailment can be compared; an instrument that shows good construct validity might differentiate between them; also

whether the group was measured before and after treatment. A correlation coefficient will be used to examine construct validity [23].

## Content validity

It is the degree in which the instrument includes the contents that are going to be evaluated [24].

Content validity will be measured by an empirical evaluation in order to find out if the set of items comprise a representative sample of the content pretended to be measured. This evaluation will be based on the opinion of experts committee and authors who have been responsible of the translation and validation steps, pilot study results and analysis of the cognitive interviews [27].

## Discriminative validity

This tests the degree in which a questionnaire can show differences between two groups of patients19. A Differential Item Functioning (DIF) analysis will be used for it. It allows for comparison of each item in two groups with the same instrument, so the results between English and Spanish versions will be contrasted [26].

## External validity

External validity reflects how generalizable the results are. This fact will permit to extrapolate from the sample to the population [28]. External validity 8 will be evaluated comparing Spanish SST with the Spanish versions of Quick Disability of the Arm, Shoulder and Hand (Quick-DASH), Numerical Pain Rating Scale (NPRS) and a global Health-Related-Quality-of-Life (HRQL) question.

## Floor and ceiling effects

The presence of floor or ceiling effects may have a negative effect on the quality of the instrument. If a group of patients scores primarily in the extremes, the responsiveness may be limited. When floor or ceiling effects are present, the psychometric properties of the questionnaire might be falsified. In cases of more than 15% of analysed data in the extreme (either in the minimum or in the maximum) floor or ceiling effects are considered to be present, respectively [18].

## Decentering

The final version of the questionnaire will need to be modified whether an item does not meet the evaluation of the psychometric properties. Original and translated versions of the questionnaire are open to changes and modifications during this decentering process, in order to reach the meaning equivalence between them [18,26].

## Feasibility

Feasibility assesses the ease with which subjects complete and researchers administer a questionnaire. Grammar and language difficulty of the SST will be assessed with the Fernandez-Huerta score [29], a Spanish analog of Flesch Reading Ease Score [30], using available word processing software (Microsoft® Word 2011). As a measure of data quality, the number of questions left blank by respondents will be analyzed.

## Discussion

The aim of this work is to describe the cross-cultural and adaptation process and validation of the Spanish-SST version. Reliability, validity and responsiveness are specific characteristics of each context, so an instrument that has demonstrated satisfactory psychometric properties in a specific population is not necessarily appropriate for others. Thus, the validation of the Spanish-SST will allow its generalization in the scientific community, better evaluation and more specific treatment aimed at the limitations shoulder-pain patients suffer from, as well as its **comparison between different countries.**

To our knowledge, SST has been translated and validated in other populations (Lithuanian, Dutch, Italian and Brazilian Portuguese) to evaluate shoulder disorders. The Lithuanian SST was translated in 2008 [31], whereas the Dutch version of the SST was published in 2012 [32] as well as the Italian version. That Italian SST has been curiously validated for the evaluation of shoulder pain and dysfunctions after neck dissection [33]. In 2014, the SST was validated for the Brazilian Portuguese speaking population3. Lithuanian, Dutch and Brazilian Portuguese versions of the SST have been validated to different shoulder conditions, such as chronic rotator cuff tears, inflammation, degenerative arthritis, etc. Lithuanian SST version includes surgery cases during the validation, but we decided to exclude them as other authors have done [31,32]. Additionally, cases of frozen shoulder and fractures have been considered exclusion criteria.

Distinct PROs have been used to assess external validity in SST validation studies. Constant-Murley (CM) Shoulder Assessment, Short Form 36 Health Survey, Oxford Shoulder Score and DASH are among them. We decided not to include CM scale as a specialist/clinician input is required to complete an objective assessment of the range of movement and power parts of the score, altering follow-up conditions or resources on research teams [34]. With regards to quality of life (QoL), a shorter and more efficient cost-benefit method will be employed: the global HRQoL question. The available current research shows a worldwide tendency to use single, global questions in population surveys to measure health status, QoL, and HRQoL [35].

The strength of this methodology is that it is likely to provide a translation that is comprehensible and generalisable to the Spanish general population. However, one weakness is that the current study does not test the translated tool's ease of understanding among individuals with cognitive difficulties or whose pain is controlled using pain medication.

## References

1. Burbank KM, Stevenson JH, Czarnecki GR, Dorfman J (2008) Chronic shoulder pain: part I. Evaluation and diagnosis. Am Fam Physician 77: 453-460.

2. Edwards S, Bell J, Blaine T (2006) Inflammation and shoulder pain--a perspective on rotator cuff disease, adhesive capsulitis, and osteoarthritis. US Orthopedics Review [Internet]. Recuperado a partir de: http://www.touchbriefings.com/download.cfm?fileID=7662

3. Neto J, Gesser R, Steglich V, Bonilauri Ferreira A, Gandhi M, et al. (2013) Validation of the Simple Shoulder Test in a Portuguese-Brazilian population. Is the latent variable structure and validation of the Simple Shoulder Test Stable across cultures? PloS One 8: e62890.

4. Schmidt S, Ferrer M, González M, González N, Valderas J, et al. (2014) Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg 23: 434-44.

5. Roy JS, MacDermid JC, Woodhouse LJ (2009) Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum 61: 623-632.

6. Lippit S, Harryman DI, Matsen FI (1993) A practical tool for evaluating function: The Simple Shoulder Test. the shoulder: A balance of mobilty

and stability. Rosemont (IL): American Academy of Orthopaedic Surgeons 501-518.

7. Salaffi F, Ciapetti A, Carotti M, Gasparini S, Filippucci E, et al. (2010) Clinical value of single versus composite provocative clinical tests in the assessment of painful shoulder. J Clin Rheumatol 16: 105-108.

8. Beaton D, Richards RR (1998) Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 7: 565-572.

9. Borgmästars N, Paavola M, Remes V, Lohman M, Vastamäki M (2010) Pain relief, motion, and function after rotator cuff repair or reconstruction may not persist after 16 years. Clin Orthop Relat Res 468: 2678-2689.

10. Roy JS, Macdermid JC, Faber KJ, Drosdowech DS, Athwal GS (2010) The simple shoulder test is responsive in assessing change following shoulder arthroplasty. J Orthop Sports Phys Ther 40: 413-421.

11. Wild D, Grove A, Martin M, Eremenco S, McElroy S, et al. (2005) 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 8: 94-104.

12. Paulsen A, Odgaard A, Overgaard S (2012) Translation, cross-cultural adaptation and validation of the Danish version of the Oxford hip score: Assessed against generic and disease-specific questionnaires. Bone Joint Res 1: 225-233.

13. Herdman M, Fox-Rushby J, Badia X (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res 7: 323-335.

14. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 25: 3186-3191.

15. Conrad F, Blair J, Tracy E (1999) Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews. FCSM Conf [Internet]. Recuperado a partir de: http://www.bls.gov/ore/abstract/st/st990240.htm

16. Varni JW, Seid M, Kurtin PS (2001) PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care 39: 800-812.

17. Nusbaum L, Natour J, Ferraz MB, Goldenberg J (2001) Translation, adaptation and validation of the Roland-Morris questionnaire--Brazil Roland-Morris. Braz J Med Biol Res 34: 203-210.

18. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60: 34-42.

19. Bullinger M, Alonso J, Apolone G, Leplège A, Sullivan M, et al. (1998) Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 51: 913-923.

20. Rouquette A, Falissard B (2011) Sample size requirements for the internal validation of psychiatric scales. Int J Methods Psychiatr Res 20: 235-249.

21. Hahn EA, Choi SW, Griffith JW, Yost KJ, Baker DW (2011) Health literacy assessment using talking touchscreen technology (Health LiTT): a new item response theory-based measure of health literacy. J Health Commun 16 Suppl 3: 150-162.

22. Osborne J, Costello A (2004) Sample size and subject to item ratio in principal components analysis. Pract Assess Res Eval 9: 8.

23. García de Yébenes Prous MA1, Rodríguez Salvanés F, Carmona Ortells L (2009) [Validation of questionnaires]. Reumatol Clin 5: 171-177.

24. Cook DA, Beckman TJ (2006) Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 119: 166.

25. Cronbach L (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16: 297-334.

26. Eremenco S, Cella D, Arnold B (2005) A comprehensive method for the translation and cross-cultural validation of health status questionnaires. Eval Health Prof 28: 212-232.

27. Allvin R, Ehnfors M, Rawal N, Svensson E, Idvall E (2009) Development of a questionnaire to measure patient-reported postoperative recovery: content validity and intra-patient reliability. J Eval Clin Pract 15: 411-419.

28. Malhotra N, Dávila Martínez J, Treviño Rosales M (2004) Investigación de mercados. Un enfoque aplicado. 4a ed. México: Pearson Educación.

29. Fernández-Huerta J (1959) Medidas sencillas de lecturabilidad. Consigna 214: 29-32.

30. Flesch R (1948) A new readability yardstick. J Appl Psychol 32: 221-233.

31. Ryliskis S, Piesina E, Kocius M, Marx R (2008) Cross-cultural adaptation and psychometric properties of the Lithuanian version of the Simple Shoulder Test. Acta Medica Litu 15: 163-168.

32. van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ (2012) Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg 21: 808-814.

33. Marchese C, Cristalli G, Pichi B, Manciocco V, Mercante G, et al. (2012) Italian cross-cultural adaptation and validation of three different scales for the evaluation of shoulder pain and dysfunction after neck dissection: University of California - Los Angeles (UCLA) Shoulder Scale, Shoulder Pain and Disability Index (SPADI) and Simple Shoulder Test (SST). Acta Otorhinolaryngol Ital 32: 12-7.

34. Levy O, Haddo O, Massoud S, Mullett H, Atoun E (2014) A patient-derived Constant-Murley score is comparable to a clinician-derived score. Clin Orthop Relat Res 472: 294-303.

35. Bowling A (2005) Just one question: If one question works, why ask several? J Epidemiol Community Health 59: 342-345.