

THE RELEVANCE OF CROSS-CULTURAL ADAPTATION AND CLINIMETRICS FOR PHYSICAL THERAPY INSTRUMENTS

MAHER CG, LATIMER J & COSTA LOP

Back Pain Research Group, The University of Sydney, Sydney, NSW - Australia

Correspondence to: Leonardo O. P. Costa, Back Pain Research Group, The University of Sydney, East Street Lidcombe, Sydney NSW Australia 2141, e-mail: lcos3060@usyd.edu.au

Received: 26/07/2007 - Revised: 28/07/2007 - Accepted: 01/08/2007

ABSTRACT

Background: Self-report outcome measures (questionnaires) are widely used by physiotherapists for measuring patient's health status or treatment outcomes. Most of these measurement tools were developed in English and their usefulness is very limited in non-English speaking countries such as Brazil. The only way to solve this problem is to properly adapt the relevant questionnaires into a target language and culture (e.g. Brazilian-Portuguese) and then test the instrument by checking its psychometric (clinimetric) characteristics. Objectives: The purpose of this paper was to present relevant issues in the process of cross-cultural adaptations and clinimetric testing for self-report outcome measurements. Advice on how to perform a cross-cultural adaptation, how to properly check the clinimetric properties, how to select a relevant questionnaire and how to evaluate the quality of an adapted questionnaire are provided. Additionally we present all Brazilian-Portuguese cross-cultural adaptations of low back pain measurements that we know of. Conclusions: There is a clear need for more effort in the field of cross-cultural adaptation and clinimetrics, without proper instruments, the management of patients from non-English speaking countries is compromised.

Key words: questionnaire; reliability; validity; psychometric properties; clinimetrics.

RESUMO

A importância da adaptação transcultural e clinimétrica para instrumentos de fisioterapia

Introdução: Questionários vem sendo amplamente utilizados por fisioterapeutas para medir a condição de saúde do paciente ou dos resultados de tratamento. A maioria desses instrumentos para avaliação foi desenvolvida em inglês, sendo seu uso bastante limitado em países que não usam o inglês como língua nativa, a exemplo do Brasil. A única forma de resolver esse problema é através de uma adaptação apropriada dos questionários relevantes para um alvo lingüístico e cultural (por exemplo, português do Brasil) e então testar suas características psicométricas (clinimétricas). Objetivo: A finalidade deste artigo foi a apresentar os tópicos relevantes no processo das adaptações transculturais de questionários e os seus respectivos testes clinimétricos. São fornecidas propostas sobre como realizar uma adaptação transcultural, como avaliar adequadamente as propriedades clinimétricas, como selecionar um questionário relevante e como avaliar a qualidade de um questionário adaptado. Além disso, são também apresentadas as adaptações conhecidas, para português do Brasil, dos questionários para avaliação de dor lombar. Conclusão: Existe uma clara necessidade de mais esforços na área de adaptação transcultural e clinimetria. Sem os instrumentos adequados, o gerenciamento no cuidado de pacientes nos países onde o inglês não é a língua nativa torna-se comprometido.

Palavras-chave: questionário; confiabilidade; validade; propriedades psicométricas; clinimetria.

INTRODUCTION

What do we mean by cross-cultural adaptation?

Physiotherapists commonly use questionnaires to assess treatment outcomes. These pen and paper tests provide a convenient way to assess how physical therapy treatment has affected outcomes such as pain, disability and health-related quality of life. Questionnaires can also be used by physiotherapists to screen for serious diseases such as cancer¹, to screen for co-morbidities that may be precautions to exercise², to estimate prognosis³ and to collect information on how their patient is thinking or feeling⁴. For a questionnaire to be useful a patients needs to be able to read and understand the text and the items need to make sense and be relevant to that person.

It would be possible to use the one questionnaire across the globe if all the people of the world spoke the same language, they felt and thought the same way and their lifestyles were much the same. Thankfully the world is a much more interesting place than this and to accommodate this variety it is necessary for researchers to adapt the original questionnaire so that it is comprehensible and relevant in the new setting. This process is called cross-cultural adaptation.

Cross-cultural adaptation is more complex than it may first seem. It is a common, though questionable, practice to use an English language questionnaire for anyone who speaks English. One problem with this approach is that the best English word to express a concept may vary between different countries. For example there are separate Australian/New Zealand, Canadian, USA and United Kingdom English language versions of the SF36 quality of life questionnaire to accommodate the different ways that the English language is used in these countries⁵. In the Australian/New Zealand and UK versions, item 9a, used to assess vitality, is "Did you feel full of life" whereas the Canadian and US versions use the wording "Did you feel full of pep". There are also differences in the use of imperial and metric units of measurement in the various SF36 versions to reflect the local customs. There are likely to be similar variations in the use of Portuguese language for the various countries where Portuguese is spoken. So where possible, Brazilian physiotherapists should use Brazilian-Portuguese questionnaires, not Portuguese versions.

Even within a country it would be simplistic to presume that a single questionnaire could be used for all people. In some countries like Canada and Belgium, there is more than one official language. Even in countries where there is only one official language, barriers to understanding may be present. For example in Australia although the official language is English there are over 200 languages spoken with ~16% of the population speaking a language other than English at home⁶. In the USA ~18% of the population⁷ speak a language other than English at home and of these about half have difficulty with English.

Even if the people of the world all spoke the same language a single assessment instrument would not be possible. This is because the people of the world do not share a common culture and lifestyle. As an illustration the Neck Disability Index was developed in North America to assess neck pain-related disability with two of the ten items assessing how neck pain limits recreation and driving⁸. The potential problem is that not all the world partakes in these activities. Similarly the Shoulder Disability Questionnaire⁹ was developed in Europe and includes the item "Because of my shoulder I have trouble putting on a coat or sweater" which is probably not relevant in tropical countries.

Why do we need cross-cultural adaptations of outcome measures?

Most questionnaires that a physiotherapist would use were originally developed in the English language. Cross cultural adaptation of existing English language questionnaires would enable comparisons of different populations and permit the exchange of information across cultural and linguistic barriers. The ready availability of adapted questionnaires would also stop the common, though undesirable practice, of excluding subjects from clinical trials on the grounds that they do not understand the language spoken in that country.

An important reason to adapt an existing questionnaire, is that it is much more efficient than developing a new one. There is substantial work involved in developing and validating a questionnaire. As an illustration the items from the Quebec back pain disability questionnaire¹⁰ were selected based upon a series of studies surveying clinical experts and patients that identified over 130 potential items. After this preliminary stage psychometric testing (also called clinimetric testing) was undertaken to identify the final 20 items that comprise the questionnaire.

How does cross-cultural adaptation differ from translation?

Translation is not the same as cross-cultural adaptation. Translation may simply involve a single person translating the questionnaire from one language to another. In contrast cross-cultural adaptation involves a team including translators, health professionals who would normally use the questionnaire and researchers with an understanding of clinimetrics. The process of cross-cultural adaptation includes initial translation, synthesis, back translation, expert committee review, pilot testing of the draft translation and psychometric evaluation. The process is designed to achieve equivalence between the original and translated versions with respect to language (semantic and idiomatic equivalence) and also equivalent experiences of daily life and meaning/ways of thinking¹¹.

It is important to understand that the goal of cross-cultural adaptation is a questionnaire equivalent to the original; which is different from aiming to produce an identical questionnaire. There can be situations where a questionnaire

item is irrelevant or not meaningful in a new setting and it makes sense to adapt or replace the item. If this is done the revised item should be chosen to reflect the construct/attribute that was being measured with the original item and it may be useful to consult the developer of the original test.

There are problems with using questionnaires that have only been translated because translation is only one of the steps involved in cross-cultural adaptation^{11,12}. Mistranslations are possible in translated questionnaires because the full process of cross-cultural adaptation includes additional steps designed to identify mistranslation. Perhaps surprisingly, overly literal translation may cause even more problems. Trying to make the translation exactly mimic the structure and ordering of words of the original text can provide a translation that is awkward to read because the syntax of the two languages may be quite different. Literal translation may also create text that makes no sense or has a very different meaning to that of the original. As an illustration there are idiomatic expressions in English language that have a meaning quite different to that normally conveyed by the component words. For example the wording from the DASS (Depression Anxiety Stress Scale) "I felt down-hearted and blue" (feeling depressed) and the item from the MHLOC (Multidimensional Health Locus of Control) "If my condition takes a turn for the worse..." (if my condition deteriorates). Another possibility is that the translated words may be understood but irrelevant in the new setting. Lastly there may also be English words that cannot be directly translated because there is no equivalent word in the target language. These potential problems are avoided with cross-cultural adaptation.

Guidelines for cross-cultural adaptation (advice for researchers)

The process of cross-cultural adaptation includes initial translation, synthesis, back translation, expert committee review, pilot testing of draft translation and psychometric evaluation¹¹⁻¹³. To illustrate the process we will presume that an English language questionnaire, originally developed in Australia, is being adapted for use in Brazil.

Guidelines for cross-cultural adaptation advise that two or more translators should independently translate the source questionnaire¹¹⁻¹³. Multiple translations permit the identification of errors or misinterpretations in translation that would not be apparent with one translation. It has been suggested that the translators should be chosen so that they have different backgrounds and include translators who do, and do not, have an understanding of the concepts being assessed in the questionnaire¹¹. It has also been advised that the translators should be translating into their mother tongue^{11,12} and preferably be both bilingual (speak both English and Portuguese) and bicultural (have lived in Australia and Brazil)¹².

Following translation from English to Portuguese the translators and an observer synthesise the multiple translations to produce a consensus Portuguese translation. Translators

blind to the original questionnaire then translate the consensus Portuguese translation back into English. As before, multiple translators who are bilingual and bicultural are preferred but for the back translation their mother tongue should be English. Beaton et al.¹¹ advise that the back translators should be naïve to the concepts involved in the original questionnaire whereas van Widenfelt et al.¹² advise that they should have some knowledge of the area. Given there are arguments for both suggestions a sensible position would be to include both types of back translators.

An expert committee then reviews the original English language questionnaire, the Brazilian translations, the consensus Brazilian translation, the English back translations and any notes taken. The committee should comprise the translators, health professionals who would normally use the questionnaire and researchers with an understanding of clinimetrics. It can also be useful to include the developers of the original questionnaire. The goal is to produce a preliminary Brazilian-Portuguese version of the questionnaire that is equivalent to the original questionnaire. This preliminary questionnaire then undergoes pilot testing with members of the target population. Subjects from a range of backgrounds are asked to complete the questionnaire and then are interviewed by the researchers. This gives an opportunity to establish that the people who will complete the questionnaire are interpreting it in the same way as the expert committee have. Lastly the new instrument needs to undergo psychometric testing because it cannot be presumed that the test properties of the English language version will apply to the adapted questionnaire.

Psychometric (Clinimetric) Properties

The aim of the process of cross cultural adaptation is to produce semantic and idiomatic equivalence between the source and the target languages/cultures. When we perform a cross-cultural adaptation of a questionnaire we assume that these procedures will retain the original psychometric properties of the questionnaire, however this assumption may not hold because there may be cultural differences between the two populations. Because it is crucial to test the psychometric properties of the adapted questionnaire in the target population after the cross-cultural adaptation procedures, a brief explanation of the psychometric concepts are described below^{14,15}.

Internal Consistency

A good questionnaire must be homogeneous; the items should be addressing different aspects of the same construct. For example the Roland Morris Disability Questionnaire¹⁶ has 24 items that address disability in low back pain patients, every single item from this questionnaire refers to a different activity but all items are related to disability in general. Most questionnaires measure a single underlying construct by using multiple items, and these items should be moderately correlated with each other, and each item should correlate with the total

scale score; these two factors form the basis of the *Internal Consistency* of the scale¹⁴. Internal consistency can be evaluated by calculating Cronbach's alpha. A low alpha value means that some items are measuring other constructs, while a very high Cronbach's alpha means that the items in the questionnaire show too much homogeneity and some items may be redundant. Current guidelines also suggest performing a factor analysis in order to confirm the internal consistency¹⁵.

Reproducibility

Reproducibility is the extent to which repeated measurement on stable subjects yields similar results¹⁷. Reproducibility comprises two related constructs: agreement and reliability. Agreement statistics describe how close the scores for repeated measures are whereas reliability statistics describe the correlation between repeated measures. With reproducible instruments clinicians and researchers will be able to draw conclusions satisfactorily, formulate theories or make claims about generalizability. It is useful to check if the measurement tool is reproducible on different occasions (intra-tester reproducibility) and with different assessors (inter-tester reproducibility). It is usually easy to interpret reliability statistics as most are expressed on a scale from 0 to 1, where zero indicates no reliability and 1 indicates perfect reliability.

Validity

In the previous paragraph we examine how reproducible a test is under different conditions. The next step is to check if the measurement tool is assessing the specific construct that it was developed, that is *validity*. We can test validity by correlating the scores of a measurement with a *gold standard* (criterion validity), however various constructs that physiotherapists assesses in clinical practice such as health-related quality of life, pain or disability have no gold standard, in this case we can test validity by correlating the scores with another tool that measures the same construct (construct validity).

Responsiveness

We can easily assume that the goal for any kind of treatment is to induce changes in patient's health status and the *responsiveness* is the ability of a questionnaire to detect clinically important changes over time, even if these changes are small¹⁵. There are two approaches to measuring responsiveness. The more typical approach is to study patients where true improvement is expected and then calculate the effect size (which is the ratio of the mean difference to the standard deviation at baseline). Another approach is to use an external criterion of true change and investigate how well the measure can discriminate between subjects who have truly improved and those who did not¹⁵.

Ceiling and floor effects

Imagine a 10-item disability questionnaire, where each item is scored from zero (meaning no disability with that item) and ten (meaning completely unable to perform that item) and the total score is the sum of the item scores. The questionnaire is unable to detect deterioration in patients who score the maximum score (100) or improvement in patients who scored the lowest score (0). *Floor or ceiling effects* are considered to be present if more than 15% of respondents achieved the lowest (floor effect) or highest possible score (ceiling effect), respectively¹⁵. Moreover ceiling and floor effects have clear implications on the reproducibility and responsiveness of the questionnaire.

How to properly test a questionnaire

Testing a questionnaire is a very time-consuming task; we first need to identify a relevant and generalizable sample for testing. A recently published guideline for evaluating measurement properties of health status questionnaires recommends that at least 100 patients are necessary to analyse all psychometric properties¹⁵. Preferably the patients should be under treatment and all patients should answer the questionnaire(s) on three occasions: the first one at baseline (in this step it is possible to calculate validity, internal consistency and ceiling and floor effects), the second testing occasion should be chosen so that the construct being measured would not be expected to have changed (for example, we do not expect an important change in pain and disability scores in chronic low back pain (LBP) patients in 24 hours); this data will be useful for analysing test-retest reproducibility. Finally the third testing occasion should be selected so that it is likely that true change has occurred in the patient's status (for example, we expect that patients with acute LBP should improve in 2-3 weeks with physical therapy treatment). This third testing occasion allows for the assessment of responsiveness of the instrument (see figure below). It is useful to include a global change scale (such as the Global Perceived Effect Scale), by using it we can establish whether the patients have changed or not over the follow up periods.

HOW TO SELECT A CROSS-CULTURAL ADAPTATION: (ADVICE FOR CLINICIANS AND RESEARCHERS WISHING TO USE A CROSS-CULTURAL ADAPTATION)

Searching for adapted questionnaires

You have decided on the patient outcomes you wish to measure and now need to find a relevant questionnaire in Brazilian-Portuguese. How do you go about doing this?

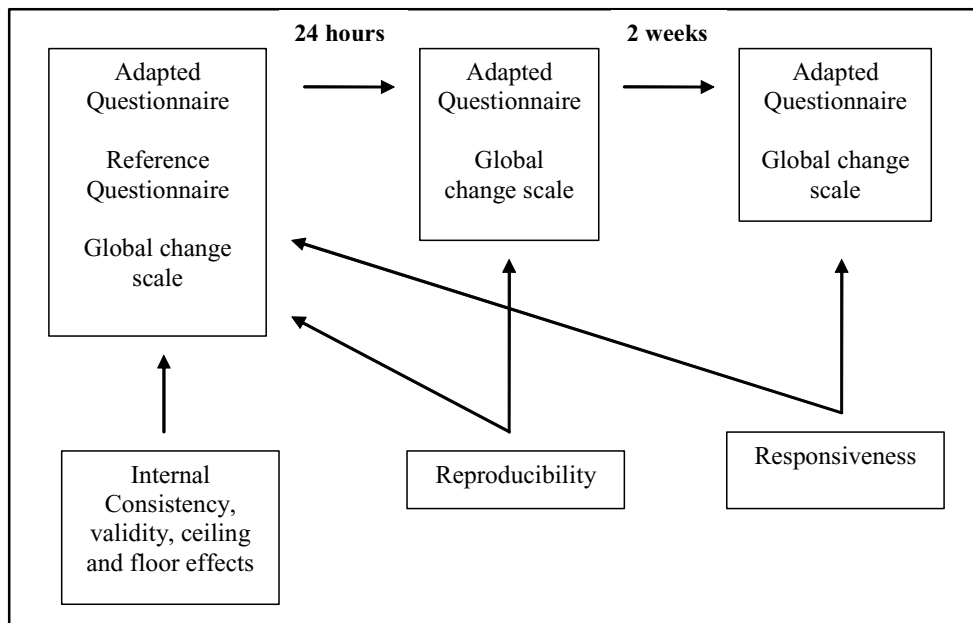


Figure 1. Example design for psychometric evaluation of an adapted questionnaire.

One strategy would be to search the internet for all relevant questionnaires that purport to measure your outcome of interest and then search for translated versions of these. If you are a researcher and subscribe to various medical databases you could follow a procedure similar to that used by Costa et al.¹⁸, in their search for international cross-cultural adaptations of self-report outcome measures for low back pain. For example, if you were interested in measuring disability in patients with low back pain you would perform two searches of the MEDLINE, EMBASE, CINAHL and LILACS databases.

The purpose of the first search would be to identify the low back pain questionnaire and could be performed by using search terms relevant to *low back pain AND questionnaire OR outcome measure*. To identify a cross-cultural adaptation, a second search would be performed where the name of the questionnaire identified in the first search would be used as the first search term (eg: *Roland Morris*) and then the terms *validation, translation, cross-cultural adaptation* plus the name of the desired language (eg: *Portuguese*) added. [The results of the search performed by Costa et al identified 40 relevant LBP self-report questionnaires of which fifteen had a cross-cultural adaptation. Only 19 of the 35 main languages that they searched for were represented in the search results, suggesting that while there are a large number of original questionnaires available, very few have been translated into other languages. Had a cross-cultural adaptation existed for each of the main languages they would have found 1400 possible adaptations (40 questionnaires x 35 languages). Clearly there is an urgent need for further cross-cultural adaptations to be conducted]¹⁸.

If you are a clinician without access to paid databases you could still search MEDLINE using PubMed which is freely available online. The PubMed URL is www.pubmed.gov.

Evaluating the quality of the located questionnaires

By now you have located a translated version of your questionnaire on the internet and are deciding whether to use it. You need to consider that there are many translated and adapted measures available on the web, some of them of very poor quality, and it is therefore important to exercise caution when deciding which measure to select. The best translated measures will be derived from peer-reviewed sources and will have followed guidelines for cross-cultural adaptation as described in the previous section. Also you need to consider whether appropriate psychometric testing of the adapted questionnaire was performed.

There are several scales available that rate the quality of psychometric testing of health questionnaires^{14,15}. The important quality criteria include satisfactory evaluation of:

- Internal consistency
- Construct validity
- Reproducibility
- Responsiveness
- Ceiling and floor effects

In order to determine the best adapted measure to use, a table is provided below with an item checklist and the criteria necessary to obtain a positive, indeterminate or negative rating (adapted from Terwee et al.¹⁵). The measure receiving the greatest number of positive ratings would be the best measure to use.

Table 1. Quality criteria for health status questionnaires.

Quality Criteria	Rating		
	+ = positive	? = indeterminate	- = negative
<i>Internal Consistency</i>	Factor analysis performed on adequate sample (≥ 100), Cronbach's alpha(s) between 0.70-0.95 for each dimension	No factor analysis OR doubtful design or method	Cronbach's alpha(s) < 0.70 or > 0.95 , despite adequate design and method
<i>Construct Validity</i>	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses	doubtful design or method (e.g. no hypotheses)	less than 75% of hypotheses were confirmed, despite adequate design and methods
<i>Reproducibility</i>	ICC or Kappa ≥ 0.70	doubtful design or method (e.g. time interval not mentioned)	ICC or Kappa < 0.70 , despite adequate design and method
<i>Responsiveness</i>	Smallest detectable change _{individual} or Smallest detectable change _{group} $<$ Minimal important change OR Minimal important change outside the limits of agreement OR Responsiveness ratio > 1.96 OR Area under the curve ≥ 0.70	doubtful design or method OR sample size < 50 OR methodological flaws	Smallest detectable change _{individual} or Smallest detectable change _{group} \geq Minimal important change OR Minimal important change equals or inside limits of agreement OR Responsiveness ratio ≤ 1.96 OR Area under the curve < 0.70 , despite adequate design and methods
<i>Ceiling and floor effects</i>	$\leq 15\%$ of the respondents achieved the highest or lowest possible scores	doubtful design or method OR sample size < 50 OR methodological flaws	$> 15\%$ of the respondents achieved the highest or lowest possible scores, despite adequate design and methods
			no information found on internal consistency
			no information found on construct validity
			no information found on reproducibility
			no information found on responsiveness
			no information found on floor and ceiling effects

The study by Costa et al.¹⁸ revealed that the quality of the psychometric evaluation of the adapted low back pain questionnaires was typically poor with most of the testing confined to the evaluation of reproducibility and construct validity of the measure. There is clearly a need for more research in this area.

Will this adapted questionnaire be useful for my patients?

In some circumstances questionnaires will have undergone cross-cultural adaptation and high quality psychometric testing but may not be useful for your patients. For instance, the items of The Roland Morris Disability Questionnaire (RMDQ) provide an insight into how a patient's low back pain is impacting on their experience of daily life. However, some of the experiences listed in this questionnaire may not be the experiences of daily life in another country or culture. Consider a Thai version of the RMDQ¹⁹ that contains the Thai translation of the English item listed below:

16. I have trouble putting on my socks (or stockings) because of the pain in my back

This question may not be relevant to many patients with low back pain living in rural areas of Thailand where putting on socks (or stockings) is not part of the daily experience. This item may also be irrelevant to people living in parts of Africa or India.

A further example of irrelevant items occurs with The Functional Rating Index (FRI) which contains items that ask about driving. There are many places in the world where such an item would not be relevant to the patient being evaluated.

In summary, careful consideration needs to be given to the items contained in the located questionnaire to determine whether they will be relevant to that patient's experience of life.

AVAILABLE BRAZILIAN-PORTUGUESE CROSS-CULTURAL ADAPTATIONS, AN EXAMPLE FROM LBP QUESTIONNAIRES

A recent systematic review performed by our research group on cross cultural adaptations for self-report outcome measures for LBP patients stated that "there is a clear need for further cross-cultural adaptation of LBP self report measures and a great attention to qualify of psychometric evaluation of adapted questionnaires"¹⁸. By the time of publication, there was just one questionnaire adapted in Brazilian-Portuguese. We have updated the data from search strategies recently and so far we have found just three questionnaires relevant to the management of LBP that are available for Brazilians.

The first adaptation in Brazilian-Portuguese was performed with the Roland Morris Disability Questionnaire (RMDQ)¹⁶, the RMDQ is a self-report questionnaire consisting of 24 items related to normal activities of daily living. The RMDQ was developed by selecting 24 relevant items from

the longer Sickness Impact Profile²⁰. The questionnaire was transformed to a condition-specific measure for back pain by adding the phrase "because of my back" to each statement. The patients are asked to circle those items, which they perceive as difficult to perform due to back pain. Each answer is scaled simply 0 or 1, thus leaving a range of scores of 0 to 24, a higher score indicating higher disability. There are two studies that tested the RMDQ in Brazilian-Portuguese, the first one²¹ tested the questionnaire in 30 Brazilian patients with LBP for reliability (ICC 0.94) and construct validity with the pain visual analogue scale (Pearson's r : 0.76 $p < 0.01$) and with pain numerical scale (0-6) (Pearson's r : 0.80 $p < 0.01$).

This questionnaire was retested by Costa et al.²² in 140 Brazilian LBP patients finding a very high level of reliability [ICC 0.95 (95%CI 0.93-0.97)] and internal consistency (Cronbach's alpha 0.92). A high correlation between the Brazil Roland Morris and the Functional Rating Index²³ was observed (Pearson's r : 0.80 $p < 0.001$) and the RMDQ correlated moderately with the Pain Numerical Rating Scale at baseline (Pearson's r : 0.55 $p < 0.001$) showing good construct validity. The authors did not find any ceiling or floor effects however the responsiveness was quite small [Effect size 0.10 (84% CI 0.04-0.16)].

The second adaptation in Brazilian-Portuguese was the Oswestry Disability Index (ODI)²⁴. The ODI measures self perceived disability in LBP patients, it is a 10-item questionnaire with each item scored on a 0-5 *Likert* scale. The ODI is scored by summing the item responses and expressing the total as percentage of the maximum score. If the patient fails to complete a section, the percentage score is adjusted. The total ODI score ranges from 0 (no disability) to 100 (maximal disability). The Brazilian-Portuguese version of the ODI²⁵ was tested in 120 LBP patients, the authors finding good internal consistency (Cronbach's alpha 0.87) and excellent reliability (ICC 0.99). There was evidence of construct validity as the ODI was moderately correlated with pain measurement (Pearson's r : 0.66) and highly correlated with the Brazilian version of the RMDQ (Pearson's r : 0.81).

The third LBP questionnaire tested in Brazil was the Functional Rating Index (FRI)²³ this questionnaire was developed to measure perceived disability in patients with back and/or neck pain²³. The FRI emphasises function while concurrently measuring the patient's opinion, attitude, and self-rating of disability. The FRI is shorter than the RMDQ and can be used for back and neck pain. The Brazilian-Portuguese version of the FRI²² was tested in 140 patients and a very high level of reliability [ICC 0.95 (95%CI 0.93-0.97)] and internal consistency (Cronbach's alpha 0.92) was found. A high correlation with the RMDQ was observed (Pearson's r : 0.80 $p < 0.001$) and the FRI correlated moderately with the Pain Numerical Rating Scale at baseline (Pearson's r : 0.67 $p < 0.001$) showing good construct validity. The authors did not find any ceiling or floor effects and the

responsiveness was small [Effect size 0.18 (84% CI 0.11-0.24)].

Directions for the future

Our recent experience of searching for international cross-cultural adaptations identified:

(i) the urgent need for more cross-cultural adaptations of useful self report outcome measures

(ii) the need for high quality psychometric evaluation of such measures

(iii) the importance of publishing the adaptation, with an additional English abstract if possible, in a journal that is indexed and therefore included in a freely available database such as *PubMed*

(iv) the need for a repository where all cross-cultural adaptations and their psychometric evaluations could be stored and then accessed by interested researchers and clinicians freely on-line. Beaton et al.¹¹ advocated this in their original publication outlining the process for cross-cultural adaptation, but to date such a repository has not been developed. It would appear most efficient if such a repository was developed and maintained by an internationally recognised, non-discipline specific professional organisation such as the Cochrane Collaboration. Such a repository would help ensure that multiple translations are not in use at the one time, and would prevent the costly and time-consuming task of replicating an already well adapted version of a questionnaire.

Acknowledgements: Chris Maher holds a fellowship funded by the National Health Medical Research Council, Australia. Leonardo Costa is a PhD supported by CAPES, Ministério da Educação, Brazil and by the Pontificia Universidade Católica de Minas Gerais, Brazil.

REFERENCES

1. Delitto A, Erhard R, Bowling R. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther.* 1995;75: 470-89.
2. Whaley M, Brubaker P, Otto R. ACSM's guidelines for exercise testing and prescription. Philadelphia: Lippincott Williams & Wilkins; 2006.
3. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain.* 1998;14:209-15.
4. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear avoidance beliefs in chronic low back pain and disability. *Pain.* 1993;52(2):157-68.
5. Trust MO. SF-36 Health survey scoring manual for english language applications: Australia/New Zealand, Canada, United Kingdom. Boston: Medical Outcomes Trust; 1994.
6. Australian Bureau of Statistics. Canberra: Australian Bureau of Statistics, 2006.
7. US Bureau of Statistics. Washington: US bureau of Statistics; 2000.
8. Vernon H, Mior S. The neck disability index: A study of reliability and validity. *J Manip Physiol Ther.* 1991;14:409-15.
9. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. *J Clin Epidemiol.* 2000;53(1):29-38.
10. Kopeck JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale: Conceptualisation and development. *J Clin Epidemiol.* 1996;49(2):151-61.
11. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-91.
12. van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev.* 2005;8(2):135-47.
13. 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;46:1417-32.
14. Streiner DL, Norman GR. Health Measurement Scales: a practical guide to their development and use. New York: Oxford Medical Publications; 2003.
15. Terwee CB, Bot SD, de Boer MR, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
16. Roland M, Morris R. A study of natural history of back pain, part 1: development of a reliable and sensitive measure of disability in low back pain. *Spine.* 1983;8:141-4.
17. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033-9.
18. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine.* 2007;32:1028-37.
19. Jirattanaphochai K, Jung S, Sumanant C, Saengnipanthkul S. Reliability of the Roland - Morris Disability Questionnaire (Thai version) for the evaluation of low back pain patients. *J Med Assoc Thai.* 2005;88(3):407-11.
20. Bergner M, Bobbitt RA, Carter WB. The sickness impact profile: development and final revision of a health status measure. *Med Care.* 1981;19:787-805.
21. Nusbaum L, Natour J, Ferraz MB, Goldenberg J. Translation, adaptation and validation of the Roland-Morris questionnaire-Brazil Roland-Morris. *Braz J Med Biol Res.* 2001;34(2): 203-10.
22. Costa LOP, Maher CG, Latimer J, Ferreira PH, Pozzi G, Ribeiro R. Psychometric Characteristics of the Brazilian-Portuguese Versions of the Functional Rating Index and the Roland Morris Disability Questionnaire. *Spine* 2007, 32(17):1902-1907.
23. Feise RJ, Menke MJ. Functional rating index - a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine.* 2001;26:78-87.
24. Fairbank JCT, Couper J, Davies JB. The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66:271-3.
25. Vigatto R, Alexandre NM, Correa Filho HR. Development of a brazilian portuguese version of the oswestry disability index: cross-cultural adaptation, reliability, and validity. *Spine.* 2007;32:481-6.