









Review Article

A good practice guide for translating and adapting hearing-related questionnaires for different languages and cultures

Deborah A. Hall^{1,2*} , Silvia Zaragoza Domingo^{3*} , Leila Z. Hamdache⁴, Vinaya Manchaiah^{5,6,7,8} , Spoorthi Thammaiah⁸ , Chris Evans⁹ , and Lena L. N. Wong¹⁰ ; On behalf of the International Collegium of Rehabilitative Audiology and TINnitus Research NETWORK

¹National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre, Nottingham, UK, ²Otology and Hearing group Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, UK, ³Neuropsychological Research Organization, Barcelona, Spain, ⁴The Carlton Academy, Nottingham, UK, ⁵Department of Speech and Hearing, Lamar State University, Beaumont, TX, USA, ⁶Department of Behavioral Sciences and Learning, The Swedish Institute for Disability Research, Linköping University, Linköping, Sweden, ⁷Department of Speech and Hearing School of Allied Health Sciences, Manipal University, Manipal, India, ⁸Department of Audiology, All India Institute of Speech and Hearing, Mysore, India, ⁹Department of Psychology, University of Roehampton, London, UK, and ¹⁰Division of Speech and Hearing Sciences, Faculty of Education, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong, China



The International Society of Audiology



Abstract

Objectives: To raise awareness and propose a good practice guide for translating and adapting any hearing-related questionnaire to be used for comparisons across populations divided by language or culture, and to encourage investigators to publish detailed steps. **Design:** From a synthesis of existing guidelines, we propose important considerations for getting started, followed by six early steps: (1) Preparation, (2, 3) Translation steps, (4) Committee Review, (5) Field testing and (6) Reviewing and finalising the translation. **Study sample:** Not applicable. **Results:** Across these six steps, 22 different items are specified for creating a questionnaire that promotes equivalence to the original by accounting for any cultural differences. Published examples illustrate how these steps have been implemented and reported, with shared experiences from the authors, members of the International Collegium of Rehabilitative Audiology and TINnitus research NETWORK. **Conclusions:** A checklist of the preferred reporting items is included to help researchers and clinicians make informed choices about conducting or omitting any items. We also recommend using the checklist to document these decisions in any resulting report or publication. Following this step-by-step guide would promote quality assurance in multinational trials and outcome evaluations but, to confirm functional equivalence, large-scale evaluation of psychometric properties should follow.

Key Words: Behavioural measures; instrumentation; psycho-social/emotional; adult or general hearing screening; tinnitus

Introduction

Patient-reported measurement instruments often refer to questionnaires that are used in a clinical setting or a clinical trial, where the responses are reported directly by the patient (or proxy) and concern some personal aspect of health, quality of life or functional status. These can be for diagnostic assessment or for evaluating the clinical efficacy of an intervention.

In the field of audiology, there is broad diversity in questionnaires used for measuring change in hearing-related problems

(Granberg et al. 2014; Barker et al. 2015; Hall et al. 2016; Plein et al. 2016). Those same reviews indicate that the majority of questionnaires have been developed in English-speaking countries; namely United States (US), United Kingdom (UK) and Australia. Therefore, when selecting patient-reported measures for diagnosis, therapeutic evaluation or audit, a first choice for researchers and clinicians in non-English-language-speaking countries is to modify an existing instrument and confirm its psychometric properties. In the rest of the article, we use the term “investigators” to encompass any clinicians and researchers who use, or might wish to use, a

*Equal first authorship.

Correspondence: Deborah A. Hall, National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham NG1 5DU, UK. E-mail: deborah.hall@nottingham.ac.uk

(Received 10 May 2017; revised 22 September 2017; accepted 28 September 2017)

ISSN 1499-2027 print/ISSN 1708-8186 online © 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1080/14992027.2017.1393565

modified instrument to target groups other than those intended by the original developer. We use the term “cross-cultural adaptation” to describe the process that considers both language issues (translation) and cultural adaptation (idiom, cultural context and lifestyle) when modifying an existing questionnaire for another geographical setting or for people in a country that has diversity in languages and cultures (Epstein, Santo, and Guillemin 2015).

Cross-cultural adaptation has clear advantages over creating a new instrument (see Beaton et al. 2000; Guillemin, Bombardier, and Beaton 1993; Wild et al. 2009). First, a multiplicity of questionnaire instruments already exist and are available ready to use. Second, many of these instruments were developed using a well-established framework. Cross-cultural adaptation is not just about translation, but also about considering the conceptual, item, semantic and operational equivalences between the source- and target-language versions and this is *essential* for enabling international research, cross-cultural comparison and meta-analysis (Herdman, Fox-Rushby, and Badia 1998). Conceptual equivalence refers to the degree to which a concept of the instrument items exist in both cultures and the meaning is the same (Herdman, Fox-Rushby, and Badia 1998). For example, “family” may be thought of as a nuclear unit in one culture (parents and offspring only) and extended (with other members) in another. Item equivalence refers to relevance of questionnaire items as measures of a particular domain. For example, an item about mowing the lawn will not be appropriate in cultures where a large proportion of the population do not own a house with a garden. Item equivalence also considers the acceptability of those questions, especially whether offensive or taboo. Semantic equivalence is concerned with the sentence structure, colloquialisms or idioms which ensure preservation of meaning. An important aspect of semantic equivalence is to ensure that the level of language used is appropriate to the end users. For accessibility, the translation should use the most widely used language variant for the country. Operational equivalence refers to the similarity of format, instructions and administration. Poor attention to these matters may compromise the overall functional equivalence, meaning that the instrument does not do “what it is supposed to do equally well in two or more cultures” (Herdman, Fox-Rushby, and Badia 1998, pp 331).

Best practice in cross-cultural adaptation is still a developing field, and numerous guidelines have been published. A systematic review identified 31 guidelines for cross-cultural research describing a similar multi-step process that aimed to promote high-quality modification of existing questionnaires, to improve the efficiency with which they are produced, and to meet regulatory body requirements (Epstein, Santo, and Guillemin 2015). Conclusions from this review highlight that guidelines share many common elements, although there is neither universal consensus among investigators on what is essential and what is optional, nor strong empirical evidence of the superiority of one method over another that might otherwise lead to a “gold standard”. Some of the 31 guidelines draw on expert recommendation by influential working parties (e.g. Guillemin, Bombardier, and Beaton 1993; Beaton et al. 2000; Wild et al. 2005; Acquadro et al. 2008; Wild et al. 2009) with authors representing specialised organisations (e.g. www.mapitrust.org/the patient-centred research company, and www.ispor.org/ the International Society for Pharmacoeconomics and Outcomes Research, ISPOR).

The International Collegium of Rehabilitative Audiology (ICRA) and TINnitus research NETwork (TINNET) represent international opinion leaders, many of whom have been involved in

the development of hearing-related questionnaires and subsequent cross-cultural adaptations (e.g. the International Outcomes Inventory for Hearing Aids, Cox, Stephens, and Kramer 2002). Discussions over the years indicate that many of our colleagues remain unaware or uncertain about what constitutes a “good” cross-cultural adaptation. Hence, this synthesis of existing recommendations is recognised to have value in promoting aspects of good practice in an original form that is accessible to the Audiology community.

The objectives of this methodological article are two-fold: first, to raise awareness and propose a good practice guide for the early steps of translating and adapting any questionnaire to be used for comparisons across populations divided by language or culture, and second to encourage publishing those details, perhaps in combination with a psychometric evaluation (not described in this guide). Our recommendations are based on common elements among well-known guidelines that have drawn in the past on expert working party recommendations for clinical trials (namely, Acquadro et al. 2008; Beaton et al. 2000; Guillemin, Bombardier, and Beaton 1993; Wild et al. 2005; Wild et al. 2009), supplemented by our own collective professional expertise. They are particularly applicable to modifications from any language to another culture or language where findings are to be interpreted or compared across countries or cultures.

Selection of the precise method will eventually depend on the competences, resources and timelines of the project. But our guidelines indicate minimum standards for any application domain, including clinical audits. For every step, we provide a short description of what is involved, with minimum standards where possible and we illustrate with examples (Supplemental File 1). Tables provide “risk indicators” to support informed decisions about the potential consequences of omitting certain steps when resources or expertise are limited. In addition, a set of editable documents are provided to guide, facilitate and boost good translation practices for future work in this field (Supplemental files 2–5). These can be modified by end users, as required.

Getting started

Scenarios requiring cross-cultural adaptation

Being faithful to an original measure is not performing a “word for word” translation but a “world for world” translation. Guillemin, Bombardier, and Beaton (1993) suggest two different contextual scenarios for when attention should be paid to cross-cultural adaptation; another country speaking another language, and new immigrants in the source country but who cannot speak the source language. The first scenario is the most common in clinical research, especially in multinational trials, when the patient-reported measure needs to be adapted into one or several languages for different countries from the one where it was created. The second scenario described by Guillemin, Bombardier, and Beaton (1993) considers new immigrants in the source country but who cannot speak the source language (e.g. Spanish speaking new immigrants arriving in the United States). This scenario also requires different language versions of the patient-reported measure to be developed, but this time used within the same country.

Other scenarios can be envisaged where the same steps are required for in-country usage. One is where there are established subpopulations living in the same country or geographical area but speaking different languages. This is the case in many Asian countries such as India and China, and also in Belgium and Canada

where subpopulations can be defined by cultural practices and linguistic dialects (e.g. Thammaiah et al. 2016). Another situation exists in countries where there is an official co-existence of several languages, with individual linguistic competencies in each official language differing, mainly due to historical reasons (e.g. Welsh and English spoken in Wales, United Kingdom UK or Catalan and Spanish spoken in Catalonia, Spain).

Scoping out the selection criteria for identifying a source questionnaire

The diversity of existing instruments for measuring the impact of hearing loss and tinnitus means that investigators can choose from a number of different patient-reported measures to assess the construct of interest. Several online resources are available for searching established data systems. A good example is “HealthMeasures” (www.healthmeasures.net/); a bank of measurement instruments for assessing global, physical, mental and social health in people living with a chronic condition [see the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative]. To guide the initial selection process, the following questions can help investigators to decide the use of one existing questionnaire instrument over another. Questions address the (i) purpose; (ii) hearing-related constructs of interest; (iii) sampling of those hearing-related constructs; (iv) psychometric properties and (v) feasibility.

(i) *For what purpose will the patient-reported measure be used?* Just because a questionnaire is popular, does not necessarily mean that it is the most appropriate. For example, a questionnaire designed primarily to discriminate between patients (diagnosis or patient stratification) will likely contain items that have different psychometric properties than one designed to evaluate changes over time (monitoring treatment outcome) (Kirshner and Guyatt, 1985). Any questionnaire to be used for evaluating intervention-related effects should have supporting evidence that it is responsive to change. An example of a questionnaire primarily developed for diagnostic use is the Tinnitus Handicap Inventory (THI; Newman, Jacobson, and Spitzer 1996), while the Tinnitus Functional Index was developed with measuring the effectiveness of interventions as its main goal (TFI; Meikle et al. 2012).

(ii) *What kind of hearing-related constructs are the focus of interest?* The US Food and Drug Administration (FDA) recommends that investigators first determine whether an adequate patient-reported measure exists to assess and measure the construct of interest (US Department of Health and Human Services FDA Center for Drug Evaluation and Research, 2006). An investigator might be interested in assessing and measuring a broad concept such as hearing disability. But disability is related to a number of discrete aspects of hearing problems such as impact on listening to speech in noise, impact on listening enjoyment, impact on social participation, etc. So, an investigator might be justified in assessing and measuring a single-domain concept instead. Few questionnaires in Audiology seem to focus on measuring a single-domain concept. Instead, most have a multidimensional structure with items assessing and measuring different concepts and combining item scores to provide a global composite score. One example is the Speech, Spatial and Qualities of Hearing Scale (SSQ, Gatehouse and Noble, 2004) which assesses three domains of hearing disability: speech communication, spatial hearing and qualities of listening. Investigators would be advised to consider the FDA guidance that a complex multidimensional claim about the clinical

efficacy of an intervention cannot be substantiated by questionnaires that do not cover component domain concepts.

(iii) *Are those constructs and how they are sampled comparable across source and target countries?* Investigators should be reassured that the concept of interest (and any associated subscale domains) is both conceptually relevant to and equivalent across source and target countries where the questionnaire will be used. This scenario is likely to be true for hearing-related conditions, but investigators should remain sensitive to the fact that the actual domains of hearing loss impact can differ across cultures. When considering whether a questionnaire should be chosen or whether any of its items need to be culturally adapted, investigators should first compare the lifestyle and listening environments between the target and the source populations (e.g. degree of urbanisation, population density, common leisure activities, religious activities, household composition, type and level of noise, etc.). If substantial proportions of the source-language questionnaire contain subscales or items which are not relevant or acceptable, then a different source questionnaire should probably be identified at the outset. As a general piece of advice, investigators should choose questionnaires that require few item changes and should avoid making excessive claims about the generalisation of a universal version without first testing it out in the field.

(iv) *Have adequate psychometric properties been demonstrated?* Properties include construct validity (the extent to which the questionnaire measures what it is supposed to measure) and reliability (the degree of measurement precision). Depending on the purpose of the questionnaire, discriminability (the degree to which the questionnaire is able to discriminate between individuals) or responsiveness (the degree to which the questionnaire is sensitive to treatment-related change) are also important. The source-language questionnaire must demonstrate adequate psychometric properties (see Valderas et al. 2008; Mokkink et al. 2010; Prinsen et al. 2016).

(v) *Is the questionnaire feasible to apply?* Feasibility is an important part of the selection process. As a minimum, feasibility should consider three essential practical aspects about the application of the candidate questionnaire; time to complete, cost and comprehensibility¹. These three criteria originated from an influential set of criteria for determining the applicability of a measurement instrument in rheumatology set by the Outcome Measures in Rheumatology consensus initiative (Boers et al. 1998).

Time to complete a questionnaire is often indicated by the number of items. For example, the SSQ (Gatehouse and Noble, 2004) contains 49 items which might render it less practical for a busy clinic or a clinical trial than a shorter instrument. Cost could be the licence fee for copyrighted materials, although often reduced tariffs are offered by authors for non-commercial (e.g. research) use. Some general examples of fee-based tools include the Hospital Anxiety and Depression Scale (owned by GL Assessment, Brentford, UK) and Health Utilities Index (Health Utilities Inc., Dundas, Ontario, Canada). Cost could also be the site staffing resources required for questionnaire administration and scoring. For example, for the SSQ an audiologist-administered interview is preferable to self-administration, in order to explain the meaning of the questions and to avoid any misunderstanding by the patient (Gatehouse and Noble, 2004). Comprehensibility (readability) refers to the degree to which an item is readily understood by most people. The linguistic diversity and literacy level of potential respondents should be considered. This preparation stage must scope out the variation in dialects spoken within the target population or the cultural variations across the target region. To

help developers create items that could easily be understood by the general public, Guillemin, Bombardier, and Beaton (1993) recommended using simple linguistic structure, such as avoiding colloquialisms, sentences containing two different verbs that suggest different actions and sentences containing two different situational contexts. However, we acknowledge that in a language/culture where colloquialisms are often used (e.g. UK English) completely avoiding them could make an instrument seem a bit dull and lifeless. To check the readability level, investigators might make use of formulas, such as the “Simple Measure of Gobbledygook” Readability formula or the Flesch Reading Ease formula (see www.readabilityformulas.com/). But, the use of these is untested in translating and adapting questionnaires. Feasibility might also extend to considerations of potential sensory problems or physical limitations that would affect a respondent’s ability to read or respond to the questionnaire.

Getting adequate resourcing to support the cross-cultural adaptation process

Achieving a good translation for multinational and multicultural research requires collaborative effort between qualified translators, healthcare professionals with experience of the condition and members of the target population (patients, communication partners, etc.). According to our step-by-step guide, a minimal team for a quality cross-cultural adaptation would involve one Translation Lead to manage the resources, procedural steps and documentation, at least three translators with linguistic competence in the source and target languages (two Forward Translators to create the target-language version, and one Back Translator to recreate a source version from that translated target), a linguistic expert (preferably a professional translator) on the target language, a healthcare professional with specific competence in the source and target languages, and the source-language Questionnaire Developer (if possible). If adequately resourced, a full cross-cultural adaptation process would typically take 4–12 weeks, depending on the difficulty of the materials to be translated, the number of review meetings required, and of reconciliations needed to reach full conceptual, item, semantic and operational equivalence.

A step-by-step guide

For each selected instrument, titles, introductory text, instructions for the administrator of the test, instructions for participants, questionnaire items, response scale anchors and scoring instructions are all equally crucial for cross-cultural adaptation. The remainder of the article gives a full step-by-step explanation of each step, presented in six sequential sections (summarised in Figure 1): (1) Preparation, (2) Translating the source language into the target language (forward translation), (3) Translating the target language back into the source language (back translation), (4) Committee Review, (5) Field testing and (6) Reviewing and finalising the translation.

Section 1 Preparation

ITEM 1A. IDENTIFY WHETHER A TARGET LANGUAGE VERSION ALREADY EXISTS, WHO DID THE DEVELOPMENT, WHETHER IT HAS CERTIFICATION DOCUMENTING THE PROCESS, AND HOW IT IS INTENDED TO BE USED

The process of cross-cultural adaptation is time-consuming and resource intensive and so before embarking on any project,

it is strongly advised to identify whether there already exists a translation of the questionnaire in the language and culture where it is going to be used (Wild et al. 2009) (Figure 1, Table 1). As a general rule, the copyright holder of the original questionnaire is also the copyright holder of the respective translated versions. If there is any doubt about whether a target-language version already exists, the copyright holder can usually signpost investigators to any translated versions and associated reports that can serve as a useful starting point. Preferably, the procedural steps of each translated version should be published (see Beaton et al. 2000), but this may not always be the case. For example, numerous translated versions of the Abbreviated Profile of Hearing Aid Benefit are instead archived on the institutional website (Hearing Aid Research Lab). Other translations might be in hands of the companies that sponsored the translations, specialised translations agencies or published in PhD theses or journals. There is at least one searchable database dedicated for clinical outcome assessments and their translations (see ePROVIDE™, <https://eprovide.mapi-trust.org/>).

If a same language version does already exist, then it is important to ascertain whether that existing version is adequate, and if not then to identify what cross-cultural adaptation steps have been done, and any limitations. In some fields, copyright owners may offer many translations of their staple measures and for some or all of these there may be no additional information regarding translation details, and almost always no peer-reviewed publication detailing the cross-cultural adaptation process and psychometric exploration of the translated versions. Translated measures without such information should in general be avoided, or taken as a starting point only. It is worth asking if the source-language Questionnaire Developer can provide a description of the process and a copy of a signed and dated certificate documenting the translation process. There is a possibility that such certification may be requested by the Independent Review Board (ethics committee) or a regulatory body (such as FDA). Practical guidelines about what should be contained in the certificate of translation are given in Supplemental File 2. It should typically include the credentials of the personnel involved, list the steps conducted, source language document, final version target language document and the person or organisation responsible for the final translation. If a certificate does not exist, but the investigators’ opinion is that the existing target language version is of good quality, then it is good practice for the investigator to conduct at least one independent back translation to confirm that the items are equivalent to the original version. If there are any concerns about semantic equivalence, then the existing translation could at least serve as one of the forward translations (see Item 2d). In the context of hearing-related questionnaire translations, the reporting of Item 1a has sometimes been unclear (e.g. Wrzosek et al. 2006; Aksoy, Firat, and Alpar 2007; Müller et al. 2016). Supplemental File 1 gives two good practice examples reporting confirmation that there was no existing target-language version.

ITEM 1B. EXPLORE PERMISSION TO USE AND TRANSLATE THE QUESTIONNAIRE, AND REQUIREMENTS NEEDED FOR GAINING APPROVAL OF THE TRANSLATED VERSION

Test developers should respect any copyright law and agreements that exist for the original questionnaire. Under certain circumstances (called “fair use”) the cross-cultural adaptation of a copyrighted work may not infringe copyright law. Nevertheless, investigators should carefully consider this matter before starting to translate any work without permission and in no matter what the

Section 1. Preparation

Adequate preparation can minimise the misuse of resources by reducing duplication of pre-existing questionnaires and by avoiding copyright infringement.

Early communication with the source-language Questionnaire Developer can help to define the concepts underlying questionnaire items to reduce the risk of misinterpretation in later stages of translation and to highlight issues from past experiences.

Section 2. Translating the source language into the target language

At least two bilingual translators should be briefed on the questionnaire, the health condition it measures and the translation requirements.

Each translator should produce an independent forward translation, from the original source language to the target language.

Discrepancies between the multiple independent translations should be reconciled and a single consolidated translation reported by a third independent team member.

Section 3. Translating the target language back into the source language

At least one additional bilingual translator should independently produce a back translation from the target language back to the original source language.

Discrepancies between the original source-language questionnaire and the back translation should be reported and resolved, including generating alternatives to consider in the next steps.

Section 4. Committee Review

A multi-disciplinary team including at least one bilingual translator and a clinician should review the forward and back translation reports and agree on solutions to discrepancies to reach cross-cultural equivalence in the forward translation.

If anything is changed this should undertake back translation in order to confirm equivalence.

Section 5. Field testing

Cognitive debriefing and/or pilot testing should be conducted with the target audience to ensure that they understand the questions and that the questions are culturally appropriate.

Section 6. Reviewing and finalising the translation

Final amendments to the questionnaire should use pilot data to ensure questions are correctly interpreted and culturally appropriate, before proofreading.

A translation certification is issued to confirm the processes.

A final report of the processes should be shared with the source-language Questionnaire Developer. As far as possible, the work should be disseminated to minimise duplication of translations into the same language.

Figure 1. Key messages from Sections 1–6, recapping on the important steps in the process of cross-cultural adaptation.

circumstance, it is always wise to make determined endeavours to contact the copyright owner. From our experience, we appreciate that it may not be possible to succeed in making contact, but the steps taken to do so should at least be reported as part of the translation process (e.g. requests sent with an acknowledgement of receipt). Wherever possible, written permission/approval could be granted from the source-language Questionnaire Developer, from his/her institution or from a publisher (if the questionnaire is published in a book, journal or publishing companies), depending on whoever holds the copyright and conditions of use. Sometimes

the copyright owner might specify certain expectations or requirements. For example, these could specify the minimal steps needed to be taken when producing translations, or could even refer to these published guidelines. The copyright owner might also stipulate what the role they wish to take in the process, and whether they will charge a fee for doing so. Some request to be actively involved in certain key steps (e.g. CORE System Trust, www.coreims.co.uk/), while others do not. Even in the case of copyright-free instruments, it is good practice to seek permission of study-specific use from the source-language Questionnaire Developer. It is not uncommon for

Table 1. Preparation.

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
1a	Identify whether a target language version already exists, who did the development, whether it has certification documenting the process, and how it is intended to be used	To evaluate if this version is fit for purpose	Unnecessary duplicate effort	T. Lead	5 pp.436
1b	Explore permission to use and translate the questionnaire, and requirements needed for gaining approval of the translated version	To respect copyright, where appropriate	Breach of copyright	T. Lead	4 pp.98
1c	Invite the source-language Questionnaire Developer to be involved	To clarify concepts or any ambiguities behind the items	Misinterpret concepts and items	T. Lead	1 pp.5152 pp.31883 pp.14234 pp.98
1d	Set out the key objectives for the process to follow, customised to the end users	To manage the project with a detailed plan	Objectives will be missed or the project will not run to plan	T. Lead	N/A
1e	Create template documents for recording the translation and adaptation process	To track all the translations and decisions across the process in a single document	Inefficient project with a high volume of documentation	T. Lead	N/A
1f	Develop definition of concepts for each questionnaire item	To enable cultural adaptation	Misinterpret concepts, meanings and items	T. Lead/QD	1 App.14 pp.98

Checklist of common steps in various recommended procedures for cross-culturally adapting patient-reported health status questionnaires. For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App. = Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp. = pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

the source-language Questionnaire Developer to request at least to be informed about the final version of the translation and provide approval before its use in research (e.g. International Outcome Inventory for Hearing Aids, <http://icra-audiology.org/>).

ITEM 1C. INVITE THE SOURCE-LANGUAGE QUESTIONNAIRE DEVELOPER TO BE INVOLVED

There are distinct advantages to inviting the source-language Questionnaire Developer to be involved in the cross-cultural adaptation process or for him/her to nominate a competent delegate. The source-language Questionnaire Developer can provide the most up-to-date information and materials to the team at the start of the project. This can include the latest existing version or formats of the questionnaire, manuals, training materials or any other useful documentation that would help in describing the concepts that are assessed and measured. Involvement is particularly beneficial at the Committee Review (Section 4) (e.g. Guillemin, Bombardier, and Beaton 1993; Beaton et al. 2000; Acquadro et al. 2008). The source-language Questionnaire Developer can share his/her accumulated knowledge, and can prompt the team to consider dialect variations, literacy levels, gender and culture issues, etc.

ITEM 1D. SET OUT THE KEY OBJECTIVES FOR THE PROCESS TO FOLLOW, CUSTOMISED TO THE END USERS

A set of self-reflective questions about (i) literacy, (ii) population characteristics and (iii) administration can help define key objectives.

(i) *Literacy levels that differ from the original source language population or diversity within the target population.* In some countries, populations may have a range of educational opportunities and literacy levels may vary greatly. If this is the case, one objective might be to use purposive sampling in recruiting participants to “pre-test” the translated version and to explicitly ask each of them to rephrase every item in their own way so that the investigators can be certain whether an item has been understood or not (e.g. Weinstein et al. 2015).

(ii) *Other characteristics of the target population.* Important patient characteristics, such as age and physical disability, can influence the choice of wording to handle stages of language development, format of administration to handle accessibility etc.

(iii) *Administration elements.* There are many different ways in which a questionnaire can be administered. During development, decisions are made about questionnaire format (written or video), instructions (for two adjacent response boxes “Pick which best...” or “Pick which best... Do not tick two boxes”), mode of administration (paper-pencil, computer, interaction with intelligent personal assistant etc.) and measurement methods (Visual Analogue Scale or Likert scale). However, a translation can only achieve operational equivalence when any changes in these elements do not affect the results (Herdman, Fox-Rushby, and Badia 1998).

ITEM 1E. CREATE TEMPLATE DOCUMENTS FOR RECORDING THE TRANSLATION AND ADAPTATION PROCESS

Usually each step is recorded in its own dedicated document to ensure that each translator works independently. However, it is

good practice to build a unified document describing all steps taken and decisions made. This information is of value to keep track in case of any future modifications and if external reviews or audits are performed. Supplemental File 3 is a template “reconciliation report” (as an excel spreadsheet) that can be modified for use, such as a unified document (see also Antunes et al. 2012).

ITEM 1F. DEVELOP DEFINITION OF CONCEPTS FOR EACH QUESTIONNAIRE ITEM

The aim of developing a clear definition of concepts in a questionnaire is to enhance conceptual, item and semantic equivalences between the source- and target-language versions (Wild et al. 2005). The preferred starting point is that the source-language Questionnaire Developer provides a list of concept definitions. If not then, for hearing-related questionnaires it would be good practice to involve audiologists or otologists in explaining any concepts that require definitions, because they have a deep understanding of the condition of interest and of what they want to measure. The Translation Lead may then seek consensus on a single definition of a concept with members of the project team. It would be good practice for the source-language Questionnaire Developer to approve the definition of concepts before any forward translation activity starts. The concept definitions can be recorded in the reconciliation report (see Supplemental File 3, column heading “Concept Definition”) or as a separate file. The most crucial thing is that it must be available to all translators when the translations are being produced because adherence to the concept definitions is an important factor in determining the quality of the final translation. Adding a version number to the list of concept definitions can be especially useful when changes might be introduced as the cross-cultural adaptation unfolds.

Section 2 Translating the source language into the target language (forward translation)

ITEM 2A. RECRUIT AT LEAST TWO DUAL-LANGUAGE TRANSLATORS WHOSE FIRST LANGUAGE IS THE TARGET LANGUAGE AND WHO ARE IN-COUNTRY RESIDENTS WHEREVER POSSIBLE, WITH A DIALECT THAT IS REPRESENTATIVE OF THE TARGET POPULATION

The main goal is to produce a final product that preserves the same meaning, is understood by the target population, and adequately reflect any nuances of the source or target language. There are different positions on the recruitment criteria and minimum number of Forward Translators (reviewed by Acquadro et al. 2008) (Figure 1, Table 2). For example, some guidelines recommend as few as two independent Forward Translators, but insist they are bilingual, with high proficiency in both languages. Other guidelines recommend more translators, but have less stringent restrictions on their fluency in the source language. Common to most guidelines are that the target language is the first language for all Forward Translators and that they should have lived experience of the target country/culture. Ideally for the minimum number of two translators, one of them should be a professional translator because they have a certified linguistic competency, and one should be a healthcare professional who has experience of the condition of interest. An advantage of this mix in skill sets is that individual biases are reduced, thus promoting a translation which is fit for purpose. The two translations can be compared and any discrepancies can be identified for subsequent discussion and resolution (see Item 2e).

When more than two independent Forward Translators are recruited, they can be targeted to address dialect or other subpopulation issues. A Curriculum Vitae for each translator should be reviewed and archived. Supplemental File 1 gives an example of reporting how translators were selected and Supplemental File 4 provides a checklist of all recommended documentation for the whole process.

ITEM 2B. BRIEF THE TRANSLATORS ON THE INSTRUMENT, CLINICAL CONCEPTS UNDERLYING THE HEALTH CONDITION OF INTEREST AND THE CONCEPT DEFINITIONS FOR EACH QUESTIONNAIRE ITEM

The goal of the translation is to maintain the same interpretation of meaning across cultures, and so this should be clearly explained to the Forward Translators using a single-briefing session that includes a description of the health concepts (see item 1f) and an explanation of how to use these definitions to achieve item-by-item semantic equivalence (see also Item 2c) (Beaton et al. 2000). Supplemental File 1 gives one example of how the translation brief has been reported. The Forward Translators can be given additional relevant information as contained in the reconciliation report (see Supplemental File 3, column headings “Source Language (Country and version)” and “Concept Definition”).

ITEM 2C. INSTRUCT TRANSLATORS ON THE REQUIREMENTS OF THE TRANSLATION (I.E. EQUIVALENCE, ACCESSIBILITY AND ACCEPTABILITY OF WORDING) AND PREFERRED TERMINOLOGY

An accurate translation is not about making a literal translation, but instead the instructions to the translators should be guided by the key objectives set out in Item 1d. Conceptual, item, semantic and operational equivalences have been discussed already. There may also be a preferred condition-specific terminology that reflects common usage by doctors and patients in the target country but which varies across cultures (such as the term for “tinnitus”). These words and phrases could constitute a glossary of terms that can be kept updated for future reference (Supplemental File 4).

It is good practice to provide the same instructions and information to all translators. At this stage, the translators should also be instructed about the priorities for conducting the translation; to maintain conceptual, item and semantic equivalences, and to promote everyday non-technical language. It can also be useful to instruct the translators to rate the difficulty of translation for each unit of the instrument because this information can be referred to when discrepancies are observed, discussed and reconciled. Useful preparatory activities could also include (i) instructing the translators on the condition and the symptoms or everyday complaints that the instrument aims to measure and (ii) providing supplementary materials written in the local language, such as patient leaflets published by health, charity or scientific organisations; especially if that material is bilingual. Supplemental File 1 gives two examples of the reporting instructions given to translators and the adaptations made as a result.

If they do not contain culturally sensitive information, instructions for scoring and for interpreting the scores of the questionnaire could be done at minimum with one Forward Translator and one Back Translator. Response options should be given due consideration. For terminology relating to response options in a Likert scale, such as *Not at all*, *Only a little*, *A moderate amount*, *Quite a lot*, *Very much indeed*, then existing terms may already be in common

Table 2. Translating the source language into the target language (forward translation).

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
2a	Recruit at least two dual-language translators whose first language is the target language and who are in-country residents wherever possible, with a dialect that is representative of the target population	To enable appropriate linguistic relevance and cultural representation	Questionnaire may not be linguistically and culturally appropriate	T. Lead	1 pp.511–5132 pp.31883 pp.14214 pp.98-95 pp.434-5, 437
2b	Brief the translators on the instrument, clinical concepts underlying the health condition of interest and the concept definitions for each questionnaire item	To provide conceptual equivalence from a clinical perspective	Lose original concept	T. Lead	1 pp.511, 5162 pp.31883 pp.14214 pp.99
2c	Instruct translators on the requirements of the translation (i.e. equivalence, accessibility, and acceptability of wording) and preferred terminology	To improve quality of translation	Questionnaire may differ from the source or not be useable by the target population	T. Lead	1 pp.510, 5162 pp.3187–31893 pp.1420 and 1421, 14234 pp.995 pp.431–433, 436 and 437
2d	Working independently, each translator produces a written record of forward translation	To ensure that the range of viewpoints is not compromised and to provide detailed information of the process	Translation reflects an individual personal style and process is not transparent	T	1 pp.511–523, 515 and 5162 pp.3187–31883 pp.14214 pp.995 pp.434-5
2e	Reconcile the forward translations to create a single forward translation; producing one written record with comments, when needed	To resolve discrepancies between the independent translations, to seek agreement between personal style and preferences, and to provide one version for the back translation	To have multiple possible translations and use research resources inefficiently at the next steps	T. Lead/T/Rev	1 pp.5152 pp.3187 and 31884 pp.995 pp.437

For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App. = Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp. = pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

usage in other target-language questionnaires. These could prove to be a good starting point. If, however, such information is not available, the Translation Lead should carefully assess whether the translated response options have the same interpretation as those used in the original source-language questionnaire. Of particular note, some cultures are less forthcoming about selecting response options at the extreme ends of a scale; narrowing the response range for statistical analysis. Furthermore, response options originally intended to be equidistant may not maintain equidistance in a literal translation.

ITEM 2D. WORKING INDEPENDENTLY, EACH TRANSLATOR PRODUCES A WRITTEN RECORD OF FORWARD TRANSLATION

Each translator should independently work to the brief provided in Items 2b/c to create a translation from source to target language, unit-by-unit. An example of reporting independent working can be found in Supplemental File 1. Moreover, relevant parts of the reconciliation report can be completed separately by each translator (see Supplemental File 3, column headings “Forward TR1 NEW LANGUAGE” and “Forward TR2 NEW LANGUAGE”). While optional for each Forward Translator to provide comments, such information can help to highlight any particular items that were difficult to translate or to document the decision taken for future

reference. At this stage, Forward Translators could rate the degree of difficulty in translating each unit of the instrument using a Likert scale to inform later review and reconciliation.

ITEM 2E. RECONCILE THE FORWARD TRANSLATIONS TO CREATE A SINGLE FORWARD TRANSLATION; PRODUCING ONE WRITTEN RECORD WITH COMMENTS, WHEN NEEDED

The aim of this step is to harmonise the forward translations. If there are any discrepancies among the two Forward Translators, then these need to be resolved by a Reviewer who makes an independent decision, in consultation with the translators when needed. When a literal translation of the word/phrase is not possible, attempts should be made to consider closest possible meaning, using the concept definition as a guide. If a consensus is not reached, then the Translation Lead could decide the final version based on the input from source-language Questionnaire Developer or by consulting a new translator or by consulting other people who do not necessarily speak the source language but who can nevertheless comment on any differences between the forward translations. The Translation Lead should not suggest new translation options because this would compromise the process. Upon review of each translation, the person in charge of the reconciliation should highlight unit-by-unit in the forward translation, each section of the text that is discrepant

Table 3. Translating the target language back to the source language (back translation).

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
3a	Recruit at least one new dual-language translator who is an in-country resident with experience in target culture and whose first language is the target language	To enable appropriate cultural representation and linguistic relevance	Translation may not be linguistically and culturally appropriate	T. Lead	1 pp.512 and 13, 515 and 5162 pp.3187 and 31883 pp.14225 pp.434 and 435
3b	Working independently, each translator produces a written record of back translation	To ensure that the range of viewpoints is not compromised and to provide detailed information of the process	Translation reflects an individual personal style and process is not transparent	T	1 pp.512 and 513, 5152 pp.3187 and 31883 pp.14225 pp.434 and 435
3c	Review the back translations against the source language	To refine the translation by ensuring the semantic equivalence of the translation, to identify and address problematic items	A mistranslation or omission may be overlooked	T. Lead/T/ QD	1 pp.5153 pp.14224 pp.1005 pp.437

For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App. = Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp. = pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

to the source. The record generated during this stage should contain details about all the discrepancies identified and how they were resolved. It can be entered in the reconciliation report [Supplemental File 3, column heading “Reconciliation of Forward Translation by (Name)”] including comments if needed and this forms an important part of the documentation for the Committee Review (see Section 6). See also Supplemental File 1 which gives two reporting examples of how discrepancies were reconciled.

Section 3 Translating the target language back into the source language (back translation)

Back translation is a commonly used quality assessment tool, but it is not without controversy and there is no compelling evidence that this step enhances the target-language version (Epstein, Santo, and Guillemin 2015) (Figure 1, Table 3). Committee review and field testing may be sufficient, if done well (Epstein, Santo, and Guillemin 2015; Colina et al. 2017).

ITEM 3A. RECRUIT AT LEAST ONE NEW DUAL-LANGUAGE TRANSLATOR WHO IS AN IN-COUNTRY RESIDENT WITH EXPERIENCE OF THE TARGET CULTURE AND WHOSE FIRST LANGUAGE IS THE TARGET LANGUAGE

The main goal is the same as the forward translations; to produce a translation that reflects the same level of language as the original. One of the minimal criteria for recruitment into the role of Back Translator should be a bilingual speaker with lived experience of the target culture, even if the translator is not currently an in-country resident. Ideally this person should be a professional translator because they have the appropriate linguistic expertise. Some of the guidelines recommend two Back Translators (Acquadro et al. 2008) but not all do, and two is not so common in the commercial sector. This is why we have specified one as a minimum standard.

ITEM 3B. WORKING INDEPENDENTLY, EACH TRANSLATOR PRODUCES A WRITTEN RECORD OF BACK TRANSLATION

The reconciled version in the target language should be back translated at least once into the source language with the translator working to the brief provided in Item 2b. Again the translation should be done on a unit-by-unit basis, for all parts of the instrument. The relevant part of our template should be completed in a blinded way by the translator where they are given only the reconciled target language version (see Supplemental File 3, column headings “Back Translation”). While optional for the translator to provide comments, such information can help to highlight any particular items that were difficult to translate or to document the decision taken for future reference. See Supplemental File 1 for an example of describing the translation brief.

ITEM 3C. REVIEW THE BACK TRANSLATIONS AGAINST THE SOURCE LANGUAGE

The person in charge of this step should highlight, unit-by-unit, each section of the back translation text that is discrepant to the source. To help with the Committee Review, any discrepancies can be classified using an A–D scheme (e.g. Badia and Alonso 1994; Sanchez-Moreno et al. 2008). According to this scheme, A = items which show perfect semantic equivalence and good literal and semantic parallels between the back translated and source version; B = items which show satisfactory semantic equivalence, but have used one or two different words; C = items which preserve the meaning of the original, but without a satisfactory semantic equivalence; and D = items which have no agreement. An example of category C is “... you had much more energy than usual?” versus “... you had more energy than usual?” (Sanchez-Moreno et al. 2008). Items classified as “D” are certainly ones requiring further action. Supplemental File 1 gives a reporting example of how the back translation was reviewed.

Table 4. Committee review.

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
4a	Appoint a multi-disciplinary committee that includes at least one bilingual member (preferably a linguist) whose first language is the target language and who are in-country residents with experience of the target culture, and one healthcare professional	To provide an additional quality control step	Cross-cultural equivalence may not be achieved (or is presumed when it may not be possible)	T. Lead	1 pp.512 and 513, 515 and 5162 pp.3187 and 31883 pp.14225 pp.434 and 435
4b	Review the translation report which includes the forward and back-translations of the questionnaire, the instructions and response scale with each other and with the source, and review the previously identified problematic items. Share translation solutions, and refer difficult items to the source-language Questionnaire Developer	To detect and deal with any translation discrepancies between the language versions and to reach a consensus	Translation includes differences between language versions making it difficult to conduct comparisons	T. Lead/T/ Rev/ QD	1 pp.512, 5152 pp.3188 and 31893 pp.1422-34 pp.101

For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App. = Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp. = pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

Section 4 Committee review

ITEM 4A. APPOINT A MULTI-DISCIPLINARY COMMITTEE THAT INCLUDES AT LEAST ONE DUAL-LANGUAGE MEMBER (PREFERABLY A LINGUIST) WHOSE FIRST LANGUAGE IS THE TARGET LANGUAGE AND WHO IS AN IN-COUNTRY RESIDENT WITH EXPERIENCE OF THE TARGET CULTURE, AND ONE HEALTHCARE PROFESSIONAL

This step involves the appointment of an expert, multi-disciplinary Committee to compare and confirm the congruence between the forward and back translations against the source-language questionnaire and to resolve any discrepancies (Figure 1, Table 4). The Committee should preferably include members with local language expertise, in-depth knowledge of the field, and expertise with the research methodology and translation process. Hence, it is advisable to have a linguistic expert, a healthcare professional with knowledge of the content area (preferably independent from the project team to avoid conflict of interests) and all Forward and Back Translators. The Translation Lead should be in close contact with the Committee during this time. The source-language Questionnaire Developer, if proficient in the target language can also be invited to participate in the Committee Review or at least (s)he can be requested to help in clarifying differences observed (if any arise) between the source and the target versions (e.g. Wild et al. 2005). Supplemental File 1 provides an example of reporting the Committee membership.

ITEM 4B. REVIEW THE TRANSLATION REPORT WHICH INCLUDES THE FORWARD AND BACK-TRANSLATIONS OF THE QUESTIONNAIRE, THE INSTRUCTIONS AND RESPONSE SCALE WITH EACH OTHER AND WITH THE SOURCE, AND REVIEW THE PREVIOUSLY IDENTIFIED CONCEPTUALLY PROBLEMATIC ITEMS. SHARE TRANSLATION SOLUTIONS, AND REFER DIFFICULT ITEMS TO THE SOURCE-LANGUAGE QUESTIONNAIRE DEVELOPER

The task of the Committee is to examine whether all the translation units are accurate and whether they map to the original intent of the

source-language Questionnaire Developer. It is easy to see how the written records (during all the substages) are crucial to this Committee Review to make the meeting efficient (Thammaiah et al. 2016). The endpoint is to reach consensus on the first final version of the questionnaire in the target language (Acquadro et al. 2008) before going to subsequent steps. This process can sometimes highlight a problem in the source questionnaire which was not previously acknowledged such as identifying an item that is simply not culturally transferable. This is a good case where having the source-language Questionnaire Developer on-side and supportive can help to resolve the issue. To go further, any change or edit introduced to the target-language version at this stage of harmonisation needs to be back translated again, with final confirmation of conceptual, item, semantic and operational equivalences. A variety of processes have been used to achieve harmonisation (Wild et al. 2005) and so we provide three different illustrative reporting examples in Supplemental File 1.

Section 5 Field Testing

Examining feasibility is the last stage of the cross-cultural adaptation process before producing the final version of the translated questionnaire (Figure 1, Table 5). Field testing can also be used to investigate any translation alternatives where no consensus was found during the Committee review.

ITEM 5A. RECRUIT A SMALL SAMPLE OF PATIENTS FROM THE TARGET POPULATION

A purposive sampling method should be used for recruitment so that there is adequate representation from across the target population in terms of the severity of the condition of interest (e.g. hearing loss), age, gender, education, regional dialect, socio-economic status and any relevant cultural factors. There is no consensus on the desired sample size in the literature, and the sample size generally varies

Table 5. Field testing.

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
5a	Recruit a small sample of patients from the target population	To promote appropriate and diverse cultural representation	Missing or inaccurate data due to patients misunderstanding instructions, items or response options	T. Lead	1 pp.512–5162 pp.31893 pp.14244 pp.1015 pp.437
5b	Assess whether the new version meets the requirements (i.e. equivalence, accessibility and acceptability of wording) and preferred terminology	To ensure that the translation can be understood by the target population, is acceptable to them, and is equivalent (i.e. conceptual, item, semantic and operational) to the original	Missing or inaccurate data due to patients misunderstanding instructions, items or response options scale	F/P	1 pp.515 and 5162 pp.31893 pp.1422, 14244 pp.1015 pp.437

For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App.= Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp.= pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

between 5 and 50. Acquadro et al. (2008) reviewed 17 such guidelines and they found that some of the guidelines do not even specify the desired sample size. We recommend that at least eight participants contribute to the pre-testing of the translated version to ensure the original instructions, items and scoring materials are clearly expressed, but where there is regional variation then sample size might need to be as high as $n=20$. No matter what is the overall sample size, each participating group of interviewees should ideally be five to eight (Antunes et al. 2012). Groups should be conducted independently and therefore would happen at different times. For pilot testing (see Item 5b), if statistical analysis is to be conducted then a larger sample size will be required. We suggest at least $n=50$ if internal consistency is to be explored using the average correlation between the questionnaire items (Cronbach's alpha) (Terwee et al. 2007). Supplemental File 1 gives three examples of reporting who evaluated the pre-final translation.

ITEM 5B. ASSESS WHETHER THE NEW VERSION MEETS THE REQUIREMENTS (I.E. EQUIVALENCE, ACCESSIBILITY AND ACCEPTABILITY OF WORDING) AND PREFERRED TERMINOLOGY

Two approaches to test a final version with members of the target audience are: (i) cognitive debriefing and (ii) pilot testing. (i) Cognitive debriefing describes a *qualitative* pretesting process of the translated instrument in the target language, ensuring that the original instructions, items and scoring materials are clearly expressed. (ii) Pilot testing obtains *quantitative* data and (if planned) *qualitative* data from the administration of the instrument in under “real-world” conditions. Pilot testing checks how the questionnaire is being completed by people filling it in full. Both methods can be suitable, depending on the final use of the instrument. Some guidelines suggest that both can be done at the same time (e.g. Acquadro et al. 2008), while others recommend to start with cognitive debriefing (e.g. Wild et al. 2005). However, cognitive debriefing alone has practical advantages (smaller samples and less costly) and is far more likely to find deviations from conceptual, item, semantic and operational equivalences. It is important that samples selected for each stage are representative of the target population; stakeholder group (patients, communication partners, caregivers, etc.), clinical severity of the condition and sociodemographic features (age, education level, gender, ethnicity, etc.).

For (i) *cognitive debriefing*, methods include a face-to-face semi-structured interview or focus group. The aim is not to elicit numerical scores, but to explore how the participants understand the questions. Patients are often asked to complete the instrument while “thinking aloud” and explain the reason for each of their responses, following which specific questions can be asked by the interviewer (York Health Economics Consortium, 2016). Questions generally cover whether there are any difficult words or phrases, how they would explain the item in their own words, and whether they would suggest any changes to the wording to make it clearer or more acceptable. The second question (i.e. asking to paraphrase in their own words) is considered the most important part of the cognitive debriefing process because this provides insight into how the interviewees actually understand the items (and returning to ask about the titles, introductory text, instructions and response scale anchors). The answers will provide clues about how comparable the translation is to the source and may expose issues of comprehension with particular groups (e.g. by dialect or years of education, etc.). Unless the participant clearly finds this difficult or impossible, it can be useful to ask participants not only on how they understand the question but to think of people they know who have had the target problems, and people who have not, and to ask them to think whether those people would (a) answer differently and (b) perhaps even read the item differently.

For (ii) *pilot testing*, the method tends to be questionnaire completion, to explore how users interact and complete the instrument. It provides an opportunity to investigate the wording of the instructions/items/response scale, its format, size, length and to understand the time necessary for the session in the target population. Investigators can also add questions about difficulty in understanding the items or response options by including supplementary Likert scales. Pilot testing is important before proceeding to a wider evaluation of its psychometric properties or before using the translation in real clinical research.

Section 6 Reviewing and finalising the translation

ITEM 6A. REVIEW THE RESULTS OF THE FIELD TESTING AND FINALISE THE TRANSLATION

The final version of the target-language questionnaire should consider incorporating the necessary modifications highlighted

Table 6. Reviewing and finalising the translation.

Item no.	Checklist item	Rationale	What are the risks of not doing this?	Who could do this?	Source
6a	Review the results of the field testing and finalise the translation	To incorporate findings of the field testing to improve performance of the translation	Translation may include words or phrases that are not familiar or not commonly used, subsequent data collected may include a high level of missing data or may be inappropriate to aggregate	T. Lead/ Rev/ QD	1 pp.515 and 5162 pp.31893 pp.14244 pp.101 and 1025 pp.437
6b	Formatting and proofreading the finalised translation	To check for minor errors which have been missed during the translation process	A final translation that contains spelling, typographical and/or other errors	PR (or Rev)	4 pp.1025 pp.437
6c	Submit all reports and other relevant documentation to the source-language Questionnaire Developer, and request for approval if applicable	To enable the source-language Questionnaire Developer to keep track of all translated versions.	The source-language Questionnaire Developer is unaware of the details of what was done	T. Lead	1 pp.5152 pp.3189
6d	Finalise and archive a report	To clearly explain the reasons for all translation choices	Work may be duplicated or translation may not be used because of inadequate reporting of methods used in development	T. Lead	1 pp.5152 pp.31894 pp.102

For brevity, the published sources are abbreviated as follows: 1 = Acquadro et al. (2008), 2 = Beaton et al. (2000), 3 = Guillemin, Bombardier, and Beaton (1993), 4 = Wild et al. (2005), 5 = Wild et al. (2009). Other abbreviations are: App.= Appendix, F = Facilitator, T. Lead = Translation Lead, P = Patient, pp.= pages, PR = Proof Reader, QD = source-language Questionnaire Developer, Rev = native-speaking in-country Reviewer, T = Translator. In-country residents specifically refer to residents of the target country.

during field testing (Figure 1, Table 6). Any problems should be taken seriously, unless recruitment included some participants different from target population (e.g. with unusually low literacy levels). Reporting examples of how the translation was finalised are given in Supplemental File 1. There appears to be no consensus about what criteria should be used for deciding whether or not to implement changes at this stage, and who should be responsible for approving those changes. It probably depends on size and representativeness of the field sample and coverage of key subpopulations (see Aquino et al. 2011). Major changes should be done only when it is absolutely necessary, should be back translated to confirm semantic equivalence and referred back to the multi-disciplinary Committee for review. An example is where items were found to be not relevant to the target culture (see Item 5b example by Weinstein et al. 2015). Our advice is to report these problematic items, so future investigators can be aware of them.

ITEM 6B. FORMATTING AND PROOFREADING THE FINALISED TRANSLATION

Investigators may perhaps take it for granted that the final version should be reviewed and proofread because we are not aware of any Audiology reporting examples. Instead we provide a “non-Audiology” reporting example (see Supplemental File 1). Proofreading for typing errors can be done by project team members. It does not need to be done by a professional service. Where the target language uses a different character set from the source language, for example, when translating from a “Latin” character set to another such as “Chinese” or “Arabic”, there may be strong preferences about fonts to be used for maximal acceptability and local expertise should guide choices and typesetting. This second step is conducted from a clerical

perspective, comparing the final version to the source-language questionnaire to ensure that the titles, introductory text, instructions for the administrator of the test, instructions for participants, question items, response scale anchors and scoring instructions are the same. Even minor differences in formatting can risk introducing substantial measurement errors and so formatting features, such as font size and styles (bold, italics, etc.) should also be the same, unless for a planned reason (e.g. a larger font for the visually impaired). Finally, unless completely impossible, the number of pages should match the source questionnaire with each question and its corresponding response option being on the same page. Usability of any formatting changes should be confirmed during pilot testing.

The finalised translation should include important supplementary information. Pages should be numbered and the document should be versioned, usually with a translation version number and date at the bottom of first page. It is also advisable to include a copyright notice for the translation and contact details of the relevant copyright holder. All supplementary information should be kept in areas of the document that do not interfere with the content of the questionnaire, such as in the document footer.

ITEM 6C. SUBMIT ALL REPORTS AND OTHER RELEVANT DOCUMENTATION TO THE SOURCE-LANGUAGE

Questionnaire developer, and request for approval if applicable. At this point, all agreements with the source-language Questionnaire Developer or the nominated delegate have to be reviewed and applied. It is a good practice that the source-language Questionnaire Developer is aware of all existing translations, even if not explicitly requested. Sharing information avoid the multiplicity of translations that can happen in research. At least, the

source-language Questionnaire Developer should receive the final translation, the certification of translation, the reconciliation reports and the concept definitions created for this purpose. A final approval and an acknowledgement of reception are always desirable, even if not required beforehand. These materials are all useful for the future.

Again, we are not aware of an Audiology example in the published literature. However, instructions on how to proceed with final translations are usually included under instructions to investigators for specific questionnaires. Two online examples are EuroQoL and HealthMeasures (see Supplemental File 1).

ITEM 6D. FINALISE AND ARCHIVE A REPORT

A written report creates a permanent record of the procedures followed, the information collected, the translation interim and final versions, and the decisions made at each stage. Supplemental File 4 contains a recommended checklist of all recommended archival documentation relating to the process of cross-cultural adaptation of patient-reported questionnaire measures. Many of these may not be published, but should be available on request. Wherever possible, it is also advisable to publish a summary of the translation process in a peer-reviewed journal for future reference. Many of the translations of questionnaires in Audiology are published as peer-reviewed journal articles, but often the details of all the different steps are not reported sufficiently well for the reader to follow exactly what was done. The items listed here (Tables 2–7) define preferred reporting standards. Supplementary File 5 itemises the preferred reporting items. We recommend that a completed list is submitted along with the manuscript to help journal reviewers locate which page of the manuscript contains a description of each individual step in the process. Similar lists exist in other areas [see the preferred reporting items in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist for systematic reviews and meta-analysis].

Besides the publication of the process, the questionnaire instrument itself can be included in the journal article. We draw attention to the possibility that this might transfer copyright ownership of that instrument to the publisher depending on the copyright agreement, with a potential unintended consequence of restricting dissemination or future modification or translation of that version of the instrument. Alternatively, the Open Access publishing with a Creative Commons public licence promotes the access and re-use of any materials included in that journal article. “Open Access” is a term indicating that the relevant work has been licenced by the copyright owner for use in some of the ways that otherwise might require their specific permission. We strongly encourage investigators to publish the translated instrument (with all titles, introductory text, instructions for the administrator of the test, instructions for participants, response scale anchors and scoring instructions) as an integral part of the journal article under a Creative Commons licence with a “No Derivatives” attribution (i.e. CC BY-ND or CC BY-NC-ND). “ND” prevents any modified versions being distributed, while “NC” prohibits commercial usage. In contrast, publishing without an “ND” attribution (i.e. CC BY or CC BY-NC) enables anyone to modify and distribute the questionnaire. We note that publishing without ND, appears to be a common practice in those Open Access articles cited in our review (e.g. Caporali et al. 2016; Rogers et al. 2016; Wrzosek et al. 2016). Putting the translated instrument out in the public domain, such as a website, yields the same loss of control over usage even if the

version is watermarked with “do not copy” (e.g. <http://harlmemphis.org/index.php?cID=130>).

Conclusions

We recognise that hearing healthcare professionals need to play a central role in good translation and adaptation of hearing-related questionnaires. Consulting with hearing professionals and members of the target populations helps to ensure that the questionnaire addresses the needs of the target population. This guideline provides step-by-step recommendations for that process. But, these are just the first essential steps because certainty about functional equivalence requires further quantitative steps to examine the psychometric soundness of the translated questionnaire instrument (e.g. Regnault and Herdman, 2015). To some readers these standards may seem laborious to follow, but they reflect the best practice and would increase expectations that the translated questionnaire instrument performs in the same way as the original. Documenting the process is equally important and we encourage investigators to publish the cross-cultural adaptation. Supplemental File 4 is the checklist of all the preferred reporting items described in this article. We recommend that investigators who are following this step-by-step guide should submit a completed version of the checklist along with the article, noting the page number corresponding to the description of each step (much like the recommended use of the PRISMA 2009 or Consolidated Standards of Reporting Trials 2010 checklists). Whenever there are difficulties complying with any of the recommendations, such as difficulties in finding bilingual speakers for the translation processes, these should be fully explained and their potential risks are carefully considered. We hope that this checklist of preferred reporting items will be adopted widely by Otolaryngology and Audiology journal editors and researchers alike.

Note

1. OMERACT use the term “interpretability” but we have chosen not to use it here because “interpretability” has another common meaning which refers to the way in which professionals might interpret the results, through instructions or training.

Acknowledgements

This work has been supported by ICRA and TINNET. Thanks to the following who submitted comments on a first draft of Tables 2–7 that were circulated for e-consultation (August 2016): Arthur Boothroyd, Judy Dubno, Sophia Kramer, Graham Naylor (ICRA members) and Berthold Langguth, Helena Caria, Christopher Cederroth, Marzena Mielczarek (TINNET members) and on the final draft of the manuscript (April 2017): Stig Arlinger, Judy Dubno, Louise Hickson, Kathy Pichora-Fuller, Gabrielle Saunders (ICRA members). Thanks also to the following members of ICRA who participated in a satellite workshop during World Congress of Audiology (Vancouver, September 2016) to discuss the scope of this article: Harvey Dillon, Judy Dubno, Kathy Pichora-Fuller, Gabrielle Saunders, Astrid van Wieringen, Sophia Kramer, Thomas Lunner, Louise Hickson, Graham Naylor, and Journal Editors-in-chief: Ross J Roeser, Brenda M Ryals. Thanks to Helen

L McCormick for translating sections of the article by Sanchez-Moreno *et al.* (2008) from Spanish into English.

Disclosure statement

The views expressed are those of the authors and not necessarily those of the societies or institutions they represent.

The work was partially supported through an independent research programme funded under the Biomedicine and Molecular Biosciences European Cooperation in Science and Technology (COST) Action framework (TINNET BM1306) which supported a Short Term Scientific Mission for DAH and SZD. LZH was supported by a Nuffield Research Placement (Nuffield Foundation). DAH is funded by the National Institute for Health Research (NIHR).

ORCID

Deborah A. Hall  <http://orcid.org/0000-0002-3804-1452>
 Silvia Zaragoza Domingo  <http://orcid.org/0000-0002-8568-4696>
 Vinaya Manchaiah  <http://orcid.org/0000-0002-1254-8407>
 Spoorthi Thammaiah  <http://orcid.org/0000-0002-1707-5901>
 Chris Evans  <http://orcid.org/0000-0002-4197-4202>
 Lena L. N. Wong  <http://orcid.org/0000-0003-2245-8592>

References

- Acquadro, C., L. Conway, A. Hareendran, and N. Aaronson. European Regulatory Issues and Quality of Life Assessment (ERIQA) Group 2008. "Literature Review of Methods to Translate Health-Related Quality of Life Questionnaires for Use in Multinational Clinical Trials." *Value in Health* 11: 509–521.
- Aksoy, S., Y. Firat, and R. Alpar. 2007. "The Tinnitus Handicap Inventory: A Study of Validity and Reliability." *The International Tinnitus Journal* 13: 94–98.
- Antunes, B., B. Daveson, C. Ramsenthaler, H. Benalia, P. L. Ferreira, C. Bausewein, and I. J. Higginson. The Palliative Care Outcome Scale. Manual for Cross-Cultural Adaptation and Psychometric Validation, London, March 2012. (<http://pos-pal.org/maix/translations.php>) [accessed 07 May 2017]
- Aquino, A. S., L. M. T. Neves, R. C. Rodrigues, and F. A. Sedin. 2011. "Translation and Cross-Cultural Adaptation of the Scoring of Patellofemoral disorders into Portuguese: Preliminary Study." *Acta Ortopédica Brasileira* 19: 273–279.
- Badia, X., and J. Alonso. 1994. "Adaptation of a Measure of Dysfunction-Related Illness: The Spanish Version of Sickness Impact Profile." *Medicina Clínica* 102: 90–95. Spanish.
- Barker, F., E. MacKenzie, L. Elliott, and S. de Lusignan. 2015. "Outcome Measurement in Adult Auditory Rehabilitation: A Scoping Review of Measures Used in Randomized Controlled Trials." *Ear Hear* 36: 567–573.
- Beaton, D. E., C. Bombardier, F. Guillemin, and M. B. Ferraz. 2000. "Guidelines for the Process of Cross-Cultural Adaptation of Self-report Measures." *Spine* 25: 3186–3191.
- Boers, M., P. Brooks, V. Strand, and P. Tugwell. 1998. "The OMERACT Filter for Outcome Measures in Rheumatology." *The Journal of Rheumatology* 25: 198–199.
- Caporali, P. F., S. A. Caporali, É. C. Bucucvic, S. Vieira Sde, Z. M. Santos, and B. M. Chiari. 2016. "Cross-Cultural Translation and Adaptation to Brazilian Portuguese of the Hearing Implant Sound Quality Index Questionnaire – (HISQUI19)." *Codas* 28: 345–354.
- Colina, S., N. Marrone, M. Ingram, and D. Sánchez. 2017. "Translation Quality Assessment in Health Research: A Functionalist Alternative to Back-Translation." *Evaluation & the Health Professions* 40: 267–293. CONSORT. Accessed 07 May 2017. <http://www.consort-statement.org/>
- Cox, R. M., D. Stephens, and S. E. Kramer. 2002. "Translations of the International Outcome Inventory for Hearing Aids (IOI-HA)." *International Journal of Audiology* 41: 3–26.
- Epstein, J., R. M. Santo, and F. Guillemin. 2015. "A Review of Guidelines for Cross-Cultural Adaptation of Questionnaires Could Not Bring Out a Consensus." *Journal of Clinical Epidemiology* 68: 435–441.
- Gatehouse, S., and W. Noble. 2004. "The Speech, Spatial and Qualities of Hearing Scale (SSQ)." *International Journal of Audiology* 43: 85–99.
- Granberg, S., J. Dahlström, C. Möller, and K. Kähäri. 2014. "The DB. ICF. Core Sets for Hearing Loss – Researcher Perspective. Part I: Systematic Review of Outcome Measures Identified in Audiological Research." *International Journal of Audiology* 53: 65–76.
- Guillemin, F., C. Bombardier, and D. Beaton. 1993. "Cross-Cultural Adaptation of Health-Related Quality of Life Measures: Literature Review and Proposed Guidelines." *Journal of Clinical Epidemiology* 46: 1417–1432.
- Hall, D. A., H. Haider, A. J. Szczeppek, P. Lau, S. Rabau, J. Jones-Diette, A. Londero, et al. 2016. "Systematic Review of Outcome Domains and Instruments Used in Clinical Trials of Tinnitus Treatments in Adults." *Trials* 17: 270.
- Hearing Aid Research Lab. Accessed 07 May 2017. <http://harlmemphis.org/index.php/clinical-applications/aphab/>
- Herdman, M., J. Fox-Rushby, and X. Badia. 1998. "A Model of Equivalence in the Cultural Adaptation of HRQoL Instruments: The Universalist Approach." *Quality of Life Research* 7: 323.
- Kirshner, B., and G. Guyatt. 1985. "A Methodological Framework for Assessing Health Indices." *Journal of Chronic Diseases* 38: 27–36.
- Meikle, M. B., J. A. Henry, S. E. Griest, B. J. Stewart, H. B. Abrams, R. McArdle, P. J. Myers, et al. 2012. The Tinnitus Functional Index: Development of a New Clinical Measure for Chronic, Intrusive Tinnitus. *Ear Hear* 33(2): 153–176.
- Mokkink, L. B., C. B. Terwee, D. L. Patrick, J. Alonso, P. W. Stratford, D. L. Knol, L. M. Bouter, and H. C. de Vet. 2010. "The COSMIN Checklist for Assessing the Methodological Quality of Studies on Measurement Properties of Health Status Measurement Instruments: An International Delphi Study." *Quality of Life Research* 19: 539–549.
- Müller, K., N. K. Edvall, E. Idrizbegovic, R. Huhn, R. Cima, V. Persson, C. Leineweber, et al. 2016. "Validation of Online Versions of Tinnitus Questionnaires Translated into Swedish." *Frontiers in Aging Neuroscience* 8: 272.
- Newman, C. W., G. P. Jacobson, and J. B. Spitzer. 1996. "Development of the Tinnitus Handicap Inventory." *Archives of Otolaryngology-Head & Neck Surgery* 122: 143–148.
- Plein, C. T., J. Harounian, E. Floyd, R. Irizarry, G. Ferzli, S. Kidwai, et al. 2016. "A Systematic Review of Eligibility and Outcomes in Tinnitus Trials. Reassessment of Tinnitus Guideline." *Otolaryngology-Head and Neck Surgery* 154: 24–32.
- Prinsen, C. A. C., S. Vohra, M. R. Rose, M. Boers, P. Tugwell, M. Clarke, P. R. Williamson, and C. B. Terwee. 2016. "How to Select Outcome Measurement Instruments for Outcomes Included in a "Core Outcome Set" – A Practical Guideline." *Trials* 17: 449.
- PRISMA. Accessed 07 May 2017. <http://www.prisma-statement.org/documents/PRISMA%202009%20checklist.pdf>
- Regnault, A., and M. Herdman. 2015. "Using Quantitative Methods Within the Universalist Model Framework to Explore the Cross-Cultural Equivalence of Patient-Reported Outcome Instruments." *Quality of Life Research* 24: 115–124.
- Rogers, K. D., M. Pilling, L. Davies, R. Belk, C. Nassimi-Green, and A. Young. 2016. "Translation, Validity and Reliability of the British Sign Language (BSL) Version of the EQ-5D-5L." *Quality of Life Research* 25: 1825–1834.

- Sanchez-Moreno, J., J. M. Villagran, J. R. Gutierrez, M. Camacho, S. Ocio, D. Palao, I. Querejeta, J. Gascon, G. Sanchez, and E. Vieta. EDHIPO (Hypomania Detection Study) Group 2008. "Adaptation and Validation of the Spanish Version of the Mood Disorder Questionnaire for the Detection of Bipolar Disorder." *Bipolar Disorders* 10: 400–412. Spanish.
- Terwee, C. B., S. D. Bot, M. R. de Boer, D. A. van der Windt, D. L. Knol, J. Dekker, L. M. Bouter, and H. C. de Vet. 2007. "Quality Criteria were Proposed for Measurement Properties of Health Status Questionnaires." *Journal of Clinical Epidemiology* 60: 34–42.
- Thammaiah, S., V. Manchaiah, E. Easwar, and R. Krishna. 2016. "Translation and Adaptation of Five English Language Self-Report Measures to South Indian Kannada Language." *Audiology Research* 6: 153.
- US Department of Health and Human Services FDA Center for Drug Evaluation and Research. 2006. FDA Center for Biologics Evaluation and Research; FDA Center for Devices and Radiological Health. "Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims: Draft Guidance." *Health and Quality of Life Outcomes* 4: 79.
- Valderas, J. M., M. Ferrer, J. Mendivil, O. Garin, L. Rajmil, M. Herdman, and J. Alonso. Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network 2008. "Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures." *Value in Health* 11: 700–708.
- Weinstein, B. A., D. Rasheedy, H. M. Taha, and F. N. Fatouh. 2015. "Cross-Cultural Adaptation of an Arabic Version of the 10-Item Hearing Handicap Inventory." *International Journal of Audiology* 54: 341–346.
- Wild, D., A. Grove, M. Martin, S. Eremenco, S. McElroy, A. Verjee-Lorenz, and P. Erikson. 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 in Health* 8: 94–104.
- Wild, D., S. Eremenco, I. Mear, M. Martin, C. Houchin, M. Gawlicki, A. Hareendran, et al. 2009. "Multinational Trials – Recommendations on the Translations Required, Approaches to Using the Same Language in Different Countries, and the Approaches to Support Pooling the Data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force Report." *Value in Health* 12: 430–440.
- Wrzosek, M., E. Szymiec, W. Klemens, P. Kotyło, W. Schlee, M. Modrzyńska, Agnieszka Lang-Malecka, Anna Preis, and J. Bulla. 2016. "Polish Translation and Validation of the Tinnitus Handicap Inventory and the Tinnitus Functional Index." *Frontiers in Psychology* 7: 1871.
- York Health Economics Consortium 2016. Cognitive Debriefing [online]. York; York Health Economics Consortium. Accessed 07 May 2017. <http://www.yhec.co.uk/glossary/cognitive-debriefing/>

Supplemental material available online