

Vertaling van meetinstrumenten

Vanwege taal- en cultuurverschillen tussen landen, is een simpele enkelvoudige vertaling van een vragenlijst niet goed genoeg. Een goede vertaling van een vragenlijst bestaat uit een proces van ten minste 2-3 onafhankelijke heenvertalingen en 2-3 onafhankelijke terugvertalingen.

Om cross-culturele vergelijkingen tussen studies mogelijk maken, is het belangrijk dat de vragenlijsten niet alleen taalkundig vergelijkbaar zijn (linguistic equivalence), maar vooral conceptueel (conceptual equivalence). Dat wil zeggen dat de vragen niet altijd exact dezelfde formulering moeten hebben, maar dat ze vooral hetzelfde moeten betekenen. Bijvoorbeeld: de Engelse vraag “do you feel down and blue”, moet niet vertaald worden met de Nederlandse vraag “voelt u zich naar beneden en blauw?” (taalkundig vergelijkbaar), maar beter met de vraag “voelt u zich neerslachtig?” (conceptueel vergelijkbaar).

Richtlijnen voor het vertalen van vragenlijsten zijn besproken in het artikel van Beaton et al, 2000 [1]. Kort samengevat is de procedure als volgt:

- 2-3 onafhankelijke heenvertalingen (bv. vanuit het Engels naar het Nederlands) door personen met de doeltaal (bv. Nederlands) als moedertaal. Eén van de vertalers heeft verstand van het concept, één vertaler heeft geen kennis van het concept dat de vragenlijst meet en ook geen medische achtergrond.
- De vertalers maken samen met een onafhankelijk persoon een gezamenlijke forward vertaling.
- 2-3 onafhankelijke terugvertalingen (bv. vanuit het Nederlands naar het Engels) door personen met de brontaal (bv. Engels) als moedertaal. Beide vertalers hebben geen kennis van het concept dat de vragenlijst meet en hebben geen medische achtergrond.
- De definitieve vertaling wordt gemaakt door een expert comité, bestaande uit de vertalers, methodologen en health care professionals, in overleg met de oorspronkelijke makers van de vragenlijst.
- Pilot studie bij 30-40 mensen uit de doelgroep waarin je het instrument wilt gaan gebruiken. Vragenlijst laten invullen en mensen achteraf interviewen over de betekenis van de items en antwoordcategorieën.

Een goede vertaling geeft geen garantie dat de vragenlijst ook betrouwbaar, valide en responsief is in de nieuwe populatie. Daarom moeten meeteigenschappen van de vertaalde vragenlijst worden onderzocht in een representatieve steekproef van de populatie waarin het vragenlijst zal worden gebruikt. Hiervoor gelden dezelfde criteria als voor het valideren van een nieuwe of bestaande vragenlijst.

Cross-culturele validiteit:

Om cross-culturele vergelijkingen tussen studies mogelijk maken, is het nodig om te onderzoeken of de scores van de vertaalde vragenlijst werkelijk vergelijkbaar zijn met scores op de oorspronkelijke vragenlijst. Hiervoor is een dataset nodig van mensen die de oorspronkelijke vragenlijst (bv. Engels) hebben ingevuld en een dataset van mensen die de vertaalde vragenlijst (bv. Nederlands) hebben ingevuld. Vervolgens kan met verschillende statistische methoden de cross-culturele validiteit worden onderzocht:

- Structural equation modelling. Dit is een vorm van confirmatieve factor analyse, waarbij gekeken wordt of de factor structuur van de vragenlijst (zeg maar de correlaties tussen de vragen onderling) vergelijkbaar is tussen de versies (zie bv. het artikel van Reed 1998 [2]).
 - Differential Item Functioning (DIF). Hierbij wordt per subschaal onderzocht of mensen met een zelfde 'niveau' (vaardigheid) op het concept, even hoog scoren op de items. Bijvoorbeeld wanneer je met een subschaal functionele status meet, verwacht je dat Engelse en Nederlandse patienten met een vergelijkbare functionele status hetzelfde scoren op de items van die schaal. Dit kan worden onderzocht met behulp van regressie-analyses (zie bv het artikel van Petersen et al 2003 [3]) of met behulp van Item Response Theory modellen (zie bv het artikel van Roorda et al, 2004 [4]).
- Voor deze analyses is het verstandig een statisticus te raadplegen, bv, Dirk Knol.

Kwaliteit van vertaling en studie naar cross-culturele validiteit

Om te beoordelen of een studie een goede vertaalprocedure heeft gevolgd of dat een studie naar cross-culturele validiteit van goede kwaliteit is, kan de box G 'cross-cultural validity' van de COSMIN checklist worden gebruikt. De COSMIN checklist is een internationale consensus-based checklist voor het beoordelen van de kwaliteit van studies naar meeteigenschappen (www.cosmin.nl). De betreffende box G uit de COSMIN checklist is op de volgende pagina's weergegeven. Deze box kan ook worden gebruikt bij de opzet van een vertaalstudie of studie naar cross-culturele validiteit om te checken of de studie volgens de geldende standaarden wordt opgezet en uitgevoerd.

COSMIN checklist - cross-cultural validity

Box G. Cross-cultural validity

<i>Design requirements</i>		yes	no	?
1	Was the percentage missing items described?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was described how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	<input type="checkbox"/>	<input type="checkbox"/>	
6	Did the translators work independently from each other?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Were items translated forward and backward?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Was adequately described how differences between the original and translated versions were resolved?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Was the translation reviewed by a committee (e.g. original developers)?	<input type="checkbox"/>	<input type="checkbox"/>	
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	<input type="checkbox"/>	<input type="checkbox"/>	
11	Was the sample used in the pre-test adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the samples similar for all characteristics except language and/or cultural background?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
14	for CTT: Was confirmatory factor analysis performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	for IRT: Was differential item function (DIF) between language groups assessed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Explanation and instructions

When evaluating cross-cultural validity, all items are applicable. When an instrument is only translated, but cross-cultural validity was not assessed, the items 4 through 11 can be used to evaluate the quality of the translation procedure.

Because of language and cultural differences, a simple translation is not sufficient. An adequate procedure contains multiple forward and backward translations with at least

two translators per step. The standards in this box are based on existing guidelines for translation and adaptation of measurement instruments, such as guidelines developed by International Quality of Life Assessment (IQOLA) [5], the MAPI Research Institute [6], or the European Organisation of Research and Treatment of Cancer (EORTC) [7].

Item 5. The characteristics and qualifications of each of the members should have been described, in terms of expertise in the languages, in the disease of the target population, and in the construct to be measured.

The specific qualifications of the translators have not been discussed in the COSMIN study. However, such specific requirements can be found in many translation guidelines. It is generally recommended that the forward translators should have the target language as their mother tongue. It is recommended that one translator has expertise on the construct to be measured, the second one being a language expert, but naïve on the topic. The back translators should have the original language as their mother tongue. They should be blind for the original version of the questionnaire. It is recommended that the back translators are both language experts and naïve to the constructs to be measured. If users of the COSMIN checklist consider the qualifications of the translators inadequate, we recommend to score item 13 with “yes”.

Item 6. To allow detection of errors, divergent interpretation or ambiguous items in the original version [8], the translators should have worked independently from each other. If only one translator was involved, item 6 should be scored “no”.

Item 7 and item 8. To further uncover mistakes in the new version, the items should have been translated forward (into the new language) and backward (back to the original language). If differences occurred between the original version, and the backward translated version, it should have been described how these differences were resolved.

Item 9. A committee should have reviewed the final translation. Preferably including the developers of the original instrument, as they know best what the items were aimed to measure. This team should be multidisciplinary, with expertise in the disease involved, and the construct to be measured, and with the involvement of members of the target population who speak the language in which the instrument was translated. These latter persons are well able to judge whether or not culturally relevant idioms are used [8].

Item 10 and item 11. A pre-test should have been performed to check the interpretation and cultural relevance of the items, and the ease of comprehension. The sample in which the translation was pre-tested should have been described in terms of age, gender, disease characteristics, and setting.

Item 12. When cross-cultural validity is assessed, the samples should be similar (e.g. in terms of age, gender, disease characteristics) except for their language.

Item 14. The preferred statistical method for assessing cross-cultural validity using CTT is confirmatory factor analysis (CFA). Based on the theoretical foundation and the factor structure of the original instrument the hypothesized factor structure can be tested using CFA.

Item 15. The preferred statistical method for assessing cross-cultural validity using IRT methods is differential item functioning (DIF) analyses [9]. DIF examines the equivalence between two versions of the same instrument. It examines whether respondents with the same level of the scale score do respond similar to a particular item. DIF can also be examined by using regression analyses (see for example [3]).

Referenties

Reference List

- [1] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25:3186-91.
- [2] Reed PJ. Medical outcomes study short form 36: testing and cross-validating a second-order factorial structure for health system employees. *Health Serv Res* 1998;33:1361-80.
- [3] Petersen MA, Groenvold M, Bjorner JB, Aaronson NK, Conroy T, Cull A, et al. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Qual Life Res* 2003;12:373-85.
- [4] Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004;63:36-42.
- [5] Bullinger M AJAGLASMW-DSGBWAANBPFSKSWJftIPG. Translating health status questionnaires and evaluating their quality: The IQOLA project approach. *Journal of Clinical Epidemiology* 1998;51:913-23.
- [6] Acquadro C, Conway K, Wolf B, Hareendran A, Mear I, Anfray C, et al. Development of a standardized classification system for the translation of Patient-Reported Outcome (PRO) measures. *PRO newsletter* 2008;39:5-7.
- [7] Koller M, Aaronson NK, Blazeby J, Bottomley A, Dewolf L, Fayers P, et al. Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur J Cancer* 2007;43:1810-20.
- [8] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993;46:1417-32.
- [9] Teresi JA, Ocepek-Welikson K, Kleinman M, Cook KF, Crane PK, Gibbons LE, et al. Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to measures of physical functioning ability and general distress. *Qual Life Res* 2007;16 Suppl 1:43-68.