ORIGINAL ARTICLE

Translation and discriminative validation of the STarT Back Screening Tool into Danish

Lars Morsø · Hanne Albert · Peter Kent ·
Claus Manniche · Jonathan Hill

Abstract

Objective The STarT Back Screening Tool (STarT) is a nine-item patient self-report questionnaire that classifies low back pain patients into low, medium or high risk of poor prognosis. When assessed by GPs, these subgroups can be used to triage patients into different evidence-based treatment pathways. The objective of this study was to translate the English version of STarT into Danish (STarT-dk) and test its discriminative validity.

Methods Translation was performed using methods recommended by best practice translation guidelines. Psychometric validation of the discriminative ability was performed using the Area Under the Curve statistic. The Area Under the Curve was calculated for seven of the nine items where reference standards were available and compared with the original English version.

Results The linguistic translation required minor semantic and layout alterations. The response options were changed from "agree/disagree" to "yes/no" for four items. No patients reported item ambiguity using the final version.

The Area Under the Curve ranged from 0.735 to 0.855 (CI95% 0.678–0.897) in a Danish cohort ($n = 311$) and 0.840 to 0.925 (CI95% 0.772–0.948) in the original English cohort ($n = 500$). On four items, the Area Under the Curve was statistically similar between the two cohorts but lower on three psychosocial sub-score items.

Conclusions The translation was linguistically accurate and the discriminative validity broadly similar, with some differences probably due to differences in severity between the cohorts and the Danish reference standard questionnaires not having been validated. Despite those differences, we believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

L. Morsø (✉) · H. Albert · P. Kent · C. Manniche
Research Department, Spine Centre of Southern Denmark,
Hospital Lillebaelt, Oestre Hougvej 55, 5500 Middelfart,
Denmark
e-mail: lars.morsoe@slb.regionsyddanmark.dk

L. Morsø · H. Albert · P. Kent · C. Manniche
Institute of Regional Health Services Research,
University of Southern Denmark, Middelfart, Denmark

J. Hill
Primary Care Musculoskeletal Research Centre,
Keele University, Keele, Staffordshire ST5 5GB, UK

Introduction

Predicting outcome from an episode of low back pain (LBP) is of interest to both the clinical and research communities [1–3]. Generic and specific questionnaires have been developed to inform that process [4–6], but there has been a need for simple tools for clinicians to use in the triage of patients in routine clinical care. The STarT Back Screening Tool (STarT) is a nine-item patient self-report questionnaire, recently validated for triage of non-specific LBP patients in primary care [7]. STarT identifies modifiable prognostic factors from the health domains of pain, activity limitation and psychosocial factors, which are risk factors for persistent non-specific LBP. STarT classifies patients into three groups: low, medium or high risk of poor prognosis, which are based on patients' symptom

complexity [7]. When assessed by GPs, these groups can be used to assist decisions about appropriate evidence-based treatment pathways. Advice and guidance is recommended for the low-risk group, referral for further treatment that focuses on physical aspects for the medium-risk group and referral for a multidimensional treatment that targets both physical and psychosocial factors for the high-risk group [7].

STarT was developed in Britain and has been translated from English into Norwegian, Dutch, French, Spanish, Welsh, Arabic and Mandarin Chinese [8], but not Danish. The objective of this study was to translate the English version of STarT into Danish and test the discriminative validity of the translated version.

Methods

This study consisted of two stages: (1) a linguistic and cultural translation and (2) a psychometric validation of discriminative ability. They were conducted at the Spine Centre of Southern Denmark, a multi-disciplinary secondary care facility. The method used was based on best practice as recommended by translation guidelines [9–11].

Linguistic and cultural translation

An overview of the phases in the linguistic and cross-cultural translation is shown in Table 1.

Phase 1: liaison with STarT developers

Contact was established with the research group at Keele University that developed STarT. This was to determine whether revisions to the English version were in progress,

Table 1 Phases in the linguistic and cultural translation

Phases		Tasks
1	Liaison with STarT developers	• Contact with the developers
		• Formation of steering committee
2	Translation	• Translation from English to Danish
3	Back translation	• Back translation from Danish to English
4	Synthesis	• Comparison of translations
5	Translation committee	• Review of translated versions
		• Reaching of consensus and development of pilot version
6	Pilot testing	• Testing in clinical setting
		• Revision of pilot version
7	Final version	• Testing of final version

to request collaboration in the translation project and access to the original validation data. A steering committee was formed consisting of two clinicians/academics whose native language was Danish and two whose native language was English. That committee included a representative of the research group at Keele.

Phase 2: translation (English to Danish)

The original STarT questionnaire was translated from English into Danish by a native Danish-speaking, professional translator. The translator was conceptually introduced to the STarT target audience and health condition and asked to take explanatory notes during the course of the translation process.

Phase 3: back translation (Danish to English)

A back translation of the Danish version was then conducted by an independent native-Danish speaking translator whose qualifications included a university degree in English. The translator was similarly conceptually introduced to the STarT target audience and health condition but the translation occurred without direct content knowledge of the original STarT version. During the translation process, explanatory notes were also taken by the translator.

Phase 4: synthesis

The content of the original and reverse-translated English versions were compared, and differences were noted. These differences were discussed with two independent native English-speaking reviewers, one of whom was a clinician and one who was not. The reviewers commented on the differences and a synthesis of these differences was created.

Phase 5: translation committee

The original English, the Danish and the reverse-translated English versions, plus the synthesis of translation differences were presented to a translation committee, who had been formed to ensure cultural relevance and conceptual equivalence. The committee consisted of seven bilingual people (clinicians/academics and lay people) and included chiropractors, physiotherapists, a surgeon and secretaries. The translation committee discussed differences in translation, whether these reflected linguistic imprecision or cultural differences, and where needed, suggested alternative wording. This process continued until consensus was reached.

Phase 6: pilot testing

A Danish pilot version of STarT was tested with 17 randomly selected LBP patients at the Spine Centre to determine the acceptability and comprehensibility of the translation. The only inclusion criteria were the presence of LBP and sufficient Danish language skills to complete the questionnaire. The pilot testing was conducted by two independent members of the translation committee, both of whom were clinicians. For each patient, their response pattern, hesitation and uncertainty while completing the pilot questionnaire were noted, along with the specific questions involved. Patients were asked about item ambiguity and difficulty. These findings were again discussed in the plenary group and the translation was adjusted, until no further hesitation or uncertainty was observed (until 'saturation').

Phase 7: final version

A second wave of testing was conducted on ten randomly selected patients using the revised questionnaire, but no further hesitation or uncertainty was observed or item ambiguity reported. This became our final Danish version of STarT, called the STarT-dk.

Psychometric validation

The discriminative validity of the STarT-dk was described using the Area Under the Curve (AUC) statistic derived from Receiver Operating Curves. The AUC was calculated for the seven items in STarT-dk where Danish versions of reference standard questionnaires were available. Each STarT-dk question was used as the 'state' variable (a yes on the screening question being the 'state') and was compared with its score on the appropriate reference standard [7]. For comparison, the AUC was also calculated for the English version using the original data used to validate the English version [7].

The reference standards were the Roland Morris Disability Questionnaire [12–14] (activity limitation), Coping Strategies Questionnaire [15, 16] (catastrophising), Tampa Scale for Kinesiophobia [17, 18], (fear of movement) and Hospital Anxiety and Depression Scale [19, 20] (anxiety and depression). All reference standards were administrated in Danish. To our knowledge Roland Morris Disability Questionnaire is the only questionnaire which has been translated into Danish using methods that meet current recommendations for cross-cultural adaptation.

Translation of the Tampa Scale for Kinesiophobia into Danish was performed at our clinical facility (forward and backward translation), but the cross-cultural adaptation is incomplete and the work is currently unpublished. The

Coping Strategies Questionnaire was sourced in a book of Scandinavian psychosocial questionnaires [21], but the quality of the translation is not reported. Four independent Danish translations of The Hospital Anxiety and Depression Scale were compared and a final Danish version was made with the purpose to approximate the original as much as possible. No further information is given on the translation and validation process [22]. These were the only available Danish language versions of the reference standards used in the English study and it was beyond the scope of the current study to validate the translations available for the reference standards.

The AUC represents the ability of the screening question to discriminate between patients with and without the symptom or sign being assessed. In statistical terms, it is sensitivity (true positive rate) divided by 1-specificity (1-true negative rate) [23]. An AUC of 1.0 is perfect discrimination and an AUC of 0.5 is discrimination no better than chance. AUC was chosen as the statistical approach because STarT is a multidimensional questionnaire consisting of one or two screening questions for each of eight underlying constructs. Therefore, multivariable analysis such as Rasch analysis [24] and other tests of internal validity are not appropriate, as they are designed for unidimensional instruments or instruments with many items measuring each construct.

Data were double-entered independently into a database (Epidata 3.1, http://www.epidata.dk "The EpiData Association" Odense, Denmark) by two research assistants. Missing values were imputed using the multiple imputation feature of PASW 18.0 (formerly SPSS) at default settings. Descriptive statistics (proportions, mean and standard deviation) were used to illustrate cohort characteristics. Differences between cohorts or subgroups were tested using Pearson Chi-square test for binomial or ordinal data and unpaired t test for continuous data. Chi-square tests were performed using Prism 5.0 (Graphpad Software Inc, La Jolla, CA, USA). All other statistical analyses were conducted using PASW 18.0 (IBM Inc., Somers, NY, USA).

Results

Linguistic and cultural translation

During phases 2 and 3, some minor linguistic differences emerged for questions 2 (pain in other body regions), 3 (walking), 4 (dressing) and 7 (catastrophising). These differences were presented to the reviewers in phase 4, who independently believed these differences to be of no consequence. For question 6, 'Worrying thoughts have been going through my mind a lot of the time' (English version)

Fig. 1 **a** The STarT*, **b** STarT-dk

was translated as 'I have been worried a lot of the time' (Danish version) as this was the closest that was linguistically possible. Linguistic difficulty in translating time factors emerged in a few questions resulting in slight potential ambiguity of these questions in Danish. In question 9 the phrase 'Overall, how bothersome …' (English version) has no equivalence in the Danish language and this was noted by the reviewers and the plenary group. Therefore, a slightly different wording closer to 'Overall, how much of an irritation…' (Danish version) was used and seemed to work in the pilot version. Initially, the translation of question 5 contained a shift from 'It's not really safe…' (English version) to 'It is not safe, really…' (Danish version) but both phrasings were tested and eventually a translation very close to the original wording was used.

The translation committee reached consensus on a pilot questionnaire that was subsequently tested in phase 6 on 17 patients. During that testing, patient hesitation and perception of item ambiguity were noted primarily for question 5 and also the response options 'Disagree' or 'Agree'. Several patients argued that some questions could not be answered in Danish as 'Disagree'/'Agree' but rather should be answered as 'No'/'Yes' and some patients suggested there was an important distinction in Danish between 'Not safe' and 'Not wise'. Therefore, the translation committee

produced a revised translation that included 'No/Yes' response options for four of the questions. The revised questionnaire was tested in phase 7 and as saturation had been achieved, no further linguistic adjustments were made. The English version of the STarT questionnaire is shown in Fig. 1a and the final Danish edition of the STarT is shown in Fig. 1b.

Psychometric validation

During this psychometric validation stage, 513 questionnaires were posted and 311 questionnaires returned—a response rate of 60.6%. The only information available on the 202 non-responders was their age and gender. Non-responders were younger (mean age 46.3 SD 14.5) that responders (mean age 51.4 SD 15.7) ($p = 0.003$ unpaired t test) and less likely to be women (non-responders 49.5% women, responders 59.5% women, $p = 0.026$ Chi-square). This resulted in a Danish cohort that was dissimilar to the English cohort in age (6.4 year difference in mean age) but almost identical in gender mix.

The participant characteristics of both the Danish and English cohorts are shown in Table 2. The prevalence of each STarT-dk classification group in the Danish cohort was 39.8% (low), 34.0% (medium) and 26.2% (high). The proportion of patients with leg pain within the last 14 days

Table 2 Participant characteristics of the Danish and the English validation samples

	Danish validation sample (n = 311)	English validation sample (n = 500)	Tests for differences between the cohorts
Female	185 (59.5%)	293 (59%)	p = 0.8912[*]
Age, mean (±SD years)	51.4 (15.7)	45 (9.7)	p = 0.0001[^]
STarT group, proportions			
Low	39.8%	47%	p = 0.788[#]
Medium	34.0%	38%	
High	26.2%	15%	
Pain intensity (scale 0–30)			
Back, mean (±SD)	17.9 (7.2)	Not comparable	p < 0.0000[*]
Leg, mean (±SD)	12.9 (9.5)	Not comparable	
Numbers with leg pain	239 (76.8%)	303 (61%)	
RMDQ (scale 0–100[§])			
Mean (±SD)	56.6 (24.1)	39.6 (25.6)	p = 0.0343[^]
Widespread pain neck/shoulders, n (%)	98 (31.5%)	276 (55%)	p = 0.0002[^]
TSK (scale 17–68[§])			
Mean (±SD)	37.5 (8.5)	39.5 (6.9)	p = 0.0003[^]
CSQ (scale 0–36[§])			
Mean (±SD)	12.9 (7.9)	10 (7.9)	p = 0.0001[^]
HADS anxiety (scale 0–20[§])			
Mean (±SD)	6.7 (4.4)	8.2 (4.5)	p = 0.0001[^]
HADS depression (scale 0–20[§])			
Mean (±SD)	4.2 (4.0)	6.7 (4.3)	p = 0.0001[^]

RMDQ Roland Morris Disability Questionnaire, *TSK* Tampa Scale of Kinesophobia, *CSQ* Coping Strategy Questionnaire (catastrophisation domain), *HADS (A)* Hospital Anxiety and Depression Scale (anxiety subscale), *HADS (D)* Hospital Anxiety and Depression Scale (depression subscale)

[*] Z test of proportions

[^] Unpaired t test

[#] Kruskal–Wallis test

[§] High scores are worse

was 79.4%, while 31.5% also reported pain in either the neck or shoulders. The mean sum score and standard deviation of the reference standard questionnaires are also reported and show that there were differences between the cohorts on most measures. This is likely to reflect the Danish cohort's being from the secondary care sector and the English cohort from the primary care sector.

Across the STarT variables in the Danish cohort, there was on average 2.19% missing data (range 5.8–0.6%). Across the reference standard variables this was 2.8% (range 10.3–1.0%). The AUC results calculated with missing values and with imputed values were compared. The largest difference in a point estimate of AUC was 0.08. Therefore, as the results calculated with missing data were almost identical to those using imputed values, only results from the original unimputed data are reported.

Identical analyses were made for seven out of nine STarT questions for both the Danish and the English

cohorts. The discriminative ability (AUC) for each of those questions in both cohorts is shown in Table 3. In the Danish cohort, AUC point estimates ranged from 0.735 to 0.855 (CI95% 0.678–0.897) and in the English cohort these point estimates ranged from 0.840 to 0.925 (CI95% 0.772–0.948). Overall, the AUC point estimates calculated for five items were similar between the two cohorts, with overlapping confidence intervals but there were differences on three psychosocial sub-score items. The discriminative ability for pain referral, pain localisation, activity limitation and fear of movement was similar. The divergence was on the anxiety, catastrophisation and depression items, as these screening questions had lower discriminative ability in the Danish cohort than in the English cohort. Therefore, correlations (Pearson r) between each individual question on each of the reference standard questionnaires for anxiety, catastrophisation and depression and the sum score of its questionnaire were calculated. This was to

Table 3 Area under Curve for each STarT question compared with its reference standard

Question on STarT	Danish Reference standard point estimate (CI95%)	English^ Reference standard point estimate (CI95%)
1. My back pain has spread down my leg(s) in the last 2 weeks	Single question: 'localisation of pain' (leg pain stated yes/no) 0.748 (CI95% 0.692–0.805)	Using a single question on current co-morbid pain sites, positive for leg pain (yes/no) 0.856 (0.784–0.927)
2. I have had pain in the shoulder or neck at some time in the last 2 weeks	Sum score of pain localisation 0.793 (CI95% 0.742–0.845)	Pain sites* 0.898 (0.842–0.955)
3. I have only walked short distances because of my back pain	RMDQ 0.846 (CI95% 0.804–0.889)	RMDQ 0.880 (0.821–0.938)
4. In the last 2 weeks, i have dressed more slowly than usual because of back pain	RMDQ 0.855 (CI95% 0.814–0.897)	RMDQ 0.846 (0.772–920)
5. It's not really safe for a person with a condition like mine to be physically active	TSK 0.775 (CI95% 0.714–0.837)	TSK 0.840 (0.770–0.908)
6. Worrying thoughts have been going through my mind a lot of the time	HADS ANX 0.837 (CI95% 0.792–0.882)	HADS ANX 0.918 (0.894–0.942)
7. I feel that my back pain is terrible and it's never going to get any better	CSQ 0.779 (CI95% 0.726–0.832)	CSQ 0.925 (0.902–0.948)
8. In general I have not enjoyed all the things I used to enjoy	HADS DEP 0.735 (CI95% 0.678–0.792)	HADS DEP 0.902 (0.876–0.929)
9. Overall, how bothersome has your back pain been in the last 2 weeks	No reference standard	No reference standard

RMDQ Roland Morris Disability Questionnaire, *TSK* Tampa Scale of Kinesophobia, *HADS* Hospital Anxiety and Depression Scale, *CSQ* Coping Strategy Questionnaire

^ Analysis performed on the external patient sample

* Estimates not comparable

examine if there was evidence that the STarT question used to screen for that construct in the English version might not be the most appropriate in this cultural setting. The results are shown in Appendix 1.

They indicate that it was uncommon that items other than those in the original STarT version displayed a stronger association with the reference standard and that when it occurred the difference in association was negligible ($r < 0.086$). How much these would vary across consecutive samples in the same language is not known, but they may be of a similar magnitude. Therefore, there must have been other reasons for this divergence in discriminative ability.

Discussion

The aim of this study was to translate the English version of STarT into Danish and test the discriminative validity of the translated version. We believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

Although the discriminative ability was similar between the two language versions for most items, there was a systematic divergence on three psychosocial items and this may have occurred for a number of reasons. The

divergence was on the anxiety, catastrophisation and depression items, as these screening questions had lower discriminative ability in the Danish cohort than in the English cohort. As one explanation for this divergence in discriminative ability might be that the screening questions chosen for the English version of STarT might not be the most appropriate for the Danish version but, as shown in Appendix 1, further analysis did not support that explanation. Another possible reason could be linguistic inaccuracies in the translation. This is not likely, as both the translators and translation committee believe the translation to be linguistically accurate.

As the three items showing divergent discriminative validity were all psychosocial constructs and as the discriminative ability was systematically lower on all three, it raises the question as to whether the divergence was the product of a cultural difference in the way Danish people answer psychosocial questions. This seems unlikely as, compared with the English cohort, the Danish cohort scored higher on some psychosocial constructs and lower on others. As the two cohorts scored differently on these psychosocial constructs, it is also possible that the divergence is due to the association between screening question and reference standard not being linear across the whole scoring range. Another reason could be the presence of inaccuracies in the Danish translation of the reference

standard questionnaires for these three psychosocial constructs. This remains a possibility, as although these translations have been performed, there are no validation studies available for those reference standards. It is also possible that some variability between samples is inevitable, even within the same language population and we do not have data to quantify that variability.

Questions of the acceptability and clinical importance of divergence have been discussed in similar contexts [25]. As the research community has not produced criterion standards to guide decisions on when divergence is clinical important, such results can only be descriptively reported.

The strengths of this study are the staged process of language translation and the direct comparison of discriminative validity with data from the original validation of the English version. A potential weakness of the study is that the translation and validation were not specifically inclusive of non-native Danish speakers as recommended by some methodologists [26, 27]. During the psychometric validation there were no restrictions on the linguistic background of the target population, but it is unknown the extent to which any indirect cross cultural validation may have occurred. The translation and psychometric validation was conducted in the secondary care sector as this was the population of convenience. Subsequent work will investigate its validity in primary care patients and determine whether the cut points in the English version are the most appropriate for the Danish version.

Conclusion

The STarT questionnaire was translated into Danish and its discriminative validity measured. The translation was judged to be linguistically accurate and the STarT-dk tool acceptable for patient use. The discriminative validity largely seemed comparable with the original English version but three psychosocial questions displayed lower discriminative validity. We suspect this is most likely to be a product of differences in severity between the cohorts and variability due to the Danish versions of the reference standard questionnaires not having been validated. Despite those differences, we believe the results show that the STarT-dk has sufficient patient acceptability and discriminative validity to be used in Denmark.

References

1. Fritz JM, Cleland JA, Childs JD (2007) Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. J Orthop Sports Phys Ther 37(6):290–302
2. Grotle M, Brox JI, Glomsrod B, Lonn JH, Vollestad NK (2007) Prognostic factors in first-time care seekers due to acute low back pain. Eur J Pain 11(3):290–298
3. Kent P, Keating JL, Leboeuf-Yde C (2010) Research methods for subgrouping low back pain. BMC Med Res Methodol 10(1):62
4. Bendebba M, Dizerega GS, Long DM (2007) The Lumbar Spine Outcomes Questionnaire: its development and psychometric properties. Spine J 7(1):118–132
5. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T et al (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 305(6846):160–164
6. Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N et al (2004) Longitudinal validation of the fear-avoidance beliefs questionnaire (FABQ) in a Swiss-German sample of low back pain patients. Eur Spine J 13(4):332–340
7. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE et al (2008) A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 59(5):632–641
8. The STarT Back Screening tool website. 2010. Ref Type: Internet Communication
9. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25(24):3186–3191
10. Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S et al (1998) Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 51(11):913–923
11. Ware JE Jr, Keller SD, Gandek B, Brazier JE, Sullivan M (1995) Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. Int J Technol Assess Health Care 11(3):525–551
12. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 8(2):141–144
13. Roland M, Fairbank J (2000) The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 25(24):3115–3124
14. Albert HB, Jensen AM, Dahl D, Rasmussen MN (2003) Criteria validation of the Roland Morris questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica. Ugeskr Laeger 165:1875–1880
15. Rosenstiel AK, Keefe FJ (1983) The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 17(1):33–44
16. Swartzman LC, Gwadry FG, Shapiro AP, Teasell RW (1994) The factor structure of the Coping Strategies Questionnaire. Pain 57(3):311–316
17. Swinkels-Meewisse EJ, Swinkels RA, Verbeek AL, Vlaeyen JW, Oostendorp RA (2003) Psychometric properties of the Tampa Scale for kinesiophobia and the fear-avoidance beliefs questionnaire in acute low back pain. Man Ther 8(1):29–36
18. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van EH (1995) Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 62(3):363–372
19. Bjelland I, Dahl AA, Haug TT, Neckelmann D (2002) The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 52(2):69–77

20. Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67(6):361–370
21. Friis-Hasché E, Elsaas P, Nielsen T (eds) (2004) Appendix 4. In: Clinical health psychology. 1st edn. Munksgaard, Copenhagen, p 435
22. Groenvold M, Fayers PM, Sprangers MA, Bjorner JB, Klee MC, Aaronson NK et al (1999) Anxiety and depression in breast cancer patients at low risk of recurrence compared with the general population: a valid comparison? J Clin Epidemiol 52:523–530
23. Kirkwood BRSJAC (1988) Measurement error: assessment and implications. In: Essential Medical Statistics. 2nd edn, Oxford: Blackwell Science Ltd., 429–446
24. Pallant JF, Tennant A (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 46(Pt 1):1–18
25. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N (2006) Danish version of the Oswestry Disability Index for patients with low back pain. Part 1: cross-cultural adaptation, reliability and validity in two different populations. Eur Spine J 15(11):1705–1716
26. Acquadro C, Conway K, Hareendran A, Aaronson N (2008) Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health 11(3):509–521
27. Altman DG, Vergouwe Y, Royston P, Moons KG (2009) Prognosis and prognostic research: validating a prognostic model. BMJ 338:b605