Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of PROMIS item banks.

Elsman, E.B.M.; Flens, G.; Beurs E. de; Roorda, L.D.; Terwee, C.B.

PLOS ONE

RESEARCH ARTICLE

Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of PROMIS item banks

Ellen B. M. Elsman[1], Gerard Flens[2], Edwin de Beurs[3,4], Leo D. Roorda[5], Caroline B. Terwee[1]*

1 Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, **2** Alliance for Quality in Mental Health Care, Utrecht, The Netherlands, **3** Arkin GGZ, Amsterdam, The Netherlands, **4** Clinical Psychology, Faculty of Social Sciences, Leiden University, Amsterdam, The Netherlands, **5** Amsterdam Rehabilitation Research Center | Reade, Amsterdam, The Netherlands

* cb.terwee@amsterdamumc.nl

OPEN ACCESS

Citation: Elsman EBM, Flens G, de Beurs E, Roorda LD, Terwee CB (2022) Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of PROMIS item banks. PLoS ONE 17(8): e0273287. https://doi.org/10.1371/journal.pone.0273287

Editor: Thiago Machado Ardenghi, Federal University of Santa Maria: Universidade Federal de Santa Maria, BRAZIL

Received: June 7, 2021

Accepted: August 5, 2022

Published: August 23, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0273287

Abstract

Introduction

The outcomes anxiety and depression are measured frequently by healthcare providers to assess the impact of a disease, but with numerous instruments. PROMIS item banks provide an opportunity for standardized measurement. Cross-cultural validity of measures and the availability of reference values are prerequisites for standardized measurement.

Methods

PROMIS Anxiety and Depression item banks were completed by 1002 representative Dutch persons. To evaluate cross-cultural validity, data from US participants in PROMIS wave 1 were used and differential item functioning (DIF) was investigated, using an iterative hybrid of logistic regression and item response theory. McFadden's pseudo R^2-change of 2% was the critical threshold. The impact of any DIF on full item banks and short forms was investigated. To obtain Dutch reference values, T-scores for anxiety and depression were calculated for the complete Dutch sample, and age-group and gender subpopulations. Thresholds corresponding to normal limits, mild, moderate and severe symptoms were computed.

Results

In both item banks, two items had DIF but with minimal impact on population level T-scores for full item banks and short forms. The Dutch general population had a T-score of 49.9 for anxiety and 49.6 for depression, similar to the T-scores of 50.0 of the US general population. T-scores for age-group and gender subpopulations were also similar to T-scores of the US general population. Thresholds for mild, moderate and severe anxiety and depression were set to 55, 60 and 70, identical to US thresholds.

Data Availability Statement: All relevant data are within the article and its Supporting Information files.

Funding: The PROMIS Health Organization is a non-profit charitable foundation and the Dutch-Flemish PROMIS National Center is a network of local members of the PHO who are developing or applying PROMIS measures in the Netherlands or Belgium. Both organizations did not provide support in the form of salaries for authors CT and LR, and did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. The author(s) received no specific funding for this work.

Competing interests: CB Terwee and LD Roorda are members of the PROMIS Health Organization and the Dutch-Flemish PROMIS National Center, which aim to improve health outcomes by developing, maintaining, improving, and encouraging the application of PROMIS in research and clinical practice. This does not alter our adherence to PLOS ONE policies on sharing data and materials. The other authors have no conflict of interest.

Conclusions

The limited number of items with DIF and its minimal impact, enables the use of standard (US) item parameters and comparisons of scores between Dutch and US populations. The Dutch reference values provide an important tool for healthcare professionals and researchers to evaluate and interpret symptoms of anxiety and depression, stimulating the uptake of PROMIS measures, and contributing to standardized outcome measurement.

Introduction

Symptoms of anxiety and depression are prevalent among patients with various conditions, such as diabetes [1], cancer [2], cardiovascular diseases cardiovascular diseases [3], and numerous mental health disorders [4, 5]. These symptoms are commonly measured by healthcare providers to assess the impact of disease and its treatment. The importance of measuring anxiety and depression is reflected in the widespread inclusion of both outcomes in Standard Sets for major medical conditions by the International Consortium for Health Outcome Measurement (ICHOM) [6]. Currently, anxiety and depression are included in 16 out of 28 Standard Sets, thereby being among the most commonly included outcomes [7].

To assess the outcomes anxiety and depression in patients or the general population, researchers and clinicians use patient-reported outcome measures (PROMs) [8–10]. Numerous PROMs assessing anxiety or depression exist, although not all of them meet the standards for reliability, validity and feasibility [11–13]. Uniformity in PROMs to measure anxiety and depression is lacking, which makes it difficult to compare their scores, and hinders benchmarking and quality of care improvements. Moreover, it is labor-intensive and costly to build in different PROMs and their scoring algorithms in electronic health records, it is difficult for healthcare providers to use different PROMs for patients with different conditions and interpret the results correctly, and it is burdensome for patients with multiple conditions to complete different PROMs measuring the same construct, which scores are not shared between healthcare providers [14–16].

To work towards standardized outcome measurement of anxiety and depression and to overcome the above mentioned challenges, the Patient-Reported Outcomes Measurement Information System (PROMIS)® initiative might provide opportunities (see [17, 18] for an overview of PROMIS and early aims and findings about the initiative). PROMIS aims to develop and maintain a state-of-the-art assessment system to measure patient-reported health with highly accurate, precise and short measures [17, 18]. The PROMIS initiative has resulted in a wide range of universal applicable (generic) item banks for use across patient populations, targeting various constructs, including item banks for measuring anxiety and depression [19–21]. PROMIS item banks can be used to create fixed questionnaires with a small number of items (also known as short forms), or used as a computerized adaptive test (CAT), which is more dynamic [22, 23]. In a CAT, items are selected from an item bank based on a persons' responses. The administration of items stops when a pre-specified criterion is met. As a result, the administration burden is reduced, with a negligible loss of precision.

The PROMIS v1.0 Anxiety item bank contains 29 items [24], whereas the v1.0 Depression item bank contains 28 items [25]. Both item banks can be applied as short forms or CAT. Fixed length short forms of 4, 6, 7 and 8 items exist for anxiety (i.e. the PROMIS Short Form v1.0 –Anxiety 4a, 6a, 7a and 8a respectively), and of 4, 6 and 8 items for depression (i.e. the PROMIS Short Form v1.0 –Depression 4a, 6a, 8a and 8b respectively) [24, 25]. Short forms

with increasing length result in more precise scores. As such, instruments intended for large scale data collection and comparisons of large groups can be short, whereas instruments intended for obtaining individual scores, diagnosing and comparing small groups should be longer. Moreover, instruments intended to monitor health status over time require more precision and thus need to be longer as well. For all these intended uses, CAT-based assessment is a good option because it combines efficiency and precision [26].

Several studies have compared PROMIS anxiety and depression instruments with legacy measures for anxiety and depression, such as the Patient Health Questionnaire, Beck Depression Inventory, General Anxiety Disorder and Centre for Epidemiological Studies Depression [27–32]. These studies conclude that PROMIS anxiety and depression instruments perform similar to these legacy measures, and can be used to screen and evaluate depression and anxiety in the general population, as well as in patient groups [27–32].

PROMIS instruments have been implemented in various institutions and health disciplines, such as orthopedics [33–35], oncology [36] and diabetes [37]. Major translation efforts have been conducted [28, 38–41], including the translation of 17 adult item banks into Dutch-Flemish [42]. The Dutch-Flemish PROMIS item banks for anxiety and depression have been validated in a representative sample of the Dutch general population as well as a clinical sample with common mental disorders [43–45], and they can be used in clinical practice and research involving the Dutch population.

In order to pursue standardized measurement of anxiety and depression and to provide contextual meaning to scores, cross-cultural validity of measures and the availability of reference values are important prerequisites. Cross-cultural validity, by means of differential item functioning (DIF) for language, has not yet been investigated for the Dutch-Flemish PROMIS item banks anxiety and depression [46]. Items should be free of DIF to ensure that the US scoring algorithm, which is the default scoring algorithm by PROMIS convention, is appropriate to use in other countries and that country-specific scores are not biased, in order to compare scores between countries. PROMIS item banks are scaled in such a way that the US general population has a mean score of 50 with a standard deviation of 10 [17]. However, some studies have shown that reference values of PROMIS scales in other countries deviate from the mean score of 50 that is obtained from the US general population [47, 48]. Therefore, this study aims to investigate DIF for language between the Netherlands and the US for the PROMIS Anxiety and Depression item banks, assess its impact, and subsequently provide reference values for these item banks for the Dutch general population.

Materials and methods

The Medical Ethical Committee of Amsterdam UMC, location VUmc, the Netherlands, confirmed that the study protocol was exempted from ethical approval according to the Dutch Medical Research in Human Subjects Act (WMO), as no experiments were conducted. The study adhered to the tenets of the Declaration of Helsinki.

Participants and procedures

Data was collected in 2014 [43, 44]. Participants were recruited from an existing internet panel of the Dutch general population by a data collection company (Desan Research Solutions). Participants needed to be representative for the Dutch general population with respect to age distribution, gender, educational level (low, middle, high), region of residence (north, east, south, west) and ethnicity (native Dutch, first- and second-generation western immigrant, first- and second-generation non-western immigrant). Representativeness of participants was compared to data from Statistics Netherlands in 2013 with maximum allowable deviations of

2.5%. Participants were asked to complete the full Dutch-Flemish PROMIS item banks Anxiety and Depression, through a web-based survey in which skipping of items was not allowed. Additionally, participants completed questions regarding their sociodemographic characteristics.

For evaluating DIF for language, data from US participants was obtained from the Health-Measures Dataverse [49], containing PROMIS wave 1 data of 21,113 participants. The calibration subsample of the anxiety and depression item banks was used [21], containing respectively 14,836 and 14,839 respondents.

To investigate how often the DIF items were included in CATs, CATs from an ongoing study in a clinical population sample of adult patients who started outpatient treatment for common mental disorders [32] were assessed.

Measures

The PROMIS Item Bank v1.0 –Anxiety consists of 29 items that assess self-reported fear (panic, fearfulness), anxious misery (dread, worry), hyperarousal (nervousness, restlessness, tension) and somatic symptoms related to arousal (dizziness, racing heart) [21, 24]. Example items include 'I felt anxious', 'I felt fearful' and 'I felt worried'. The PROMIS Item Bank v1.0 – Depression consists of 28 items that assess self-reported negative mood (guilt, sadness), views of self (worthlessness, self-criticism), social cognition (interpersonal alienation, loneliness) and decreased positive affect and engagement (loss of purpose, meaning and interest) [21, 25]. Example items include 'I felt depressed', 'I felt sad' and 'I felt lonely'. All items have a 7-day recall period and are scored on a 5-point Likert scale with response options 1 = never, 2 = rarely, 3 = sometimes, 4 = often and 5 = always. Total scores are derived from the original US IRT model (i.e. the Graded Response Model (GRM) [50]) and expressed as T-scores, with a mean of 50 and a standard deviation of 10 for the US general population [17]. Higher scores represent more anxiety/depression. In line with PROMIS convention, T-scores were calculated based on the item parameters from the original US calibration sample with expected a posteriori estimates [19]. T-scores can either be calculated by uploading item scores in the online HealthMeasures Scoring Service program, provided by the US Assessment Center [20], or by using the conversion tables in the PROMIS anxiety/depression scoring manuals to convert raw sum scores into T-scores [24, 25]. Scoring Service is the most accurate scoring method available because it uses IRT-based response pattern scoring and can handle missing data (the conversion table can only be used when all items are completed) and was therefore used for obtaining Dutch reference values in this study.

Statistical analyses

Descriptive statistics were used to summarize sociodemographic characteristics of participants. DIF analyses were conducted with an iterative hybrid of logistic regression and IRT with the lordif package [51] in R. In the logistic regression framework, three regression models were compared: model 1, in which item responses are predicted by the latent trait; model 2, in which item responses are predicted by the latent trait and group (US or NL) membership; and model 3, in which item responses are predicted by the latent trait, group membership (US or NL) and the interaction between these terms. Uniform and non-uniform DIF were assessed by comparing model 1 with model 2 and model 2 with model 3, respectively. The likelihood-ratio χ^2 test with detection criterion R2 was used to detect DIF. McFadden's pseudo R^2 was used as a measure of DIF magnitude, with a 2% change being considered as critical threshold [51, 52]. Monte Carlo simulations implemented in the lordif package (1000 replications) were performed to check for type I error inflation [51].

The impact of DIF on item and total scores was assessed by visual inspection of category response curves (CRCs) and test characteristic curves (TCCs) per group. To assess the impact of DIF on short forms and full item bank T-scores, T-scores were calculated with the original US item parameters with expected a posteriori estimates from the GRM model (obtained from HealthMeasures), which is standard practice for PROMIS measures, as well as with a hybrid set of item parameters, and subsequently compared. The hybrid set of item parameters consisted of the original (US) item parameters for the non-DIF items and rescaled Dutch item parameters for the DIF items. Dutch item parameters were obtained by fitting a GRM to the Dutch general population sample, using the mirt package [53] in R. To obtain a hybrid set of item parameters the Stocking-Lord method was used to rescale the Dutch item parameters for DIF items to the US metric [54, 55]. The equate function in the lordif package computes linear transformation constants (with DIF free items as anchor) that can be used to equate the Dutch item parameters to the scale of the US item parameters [51], while minimizing the squared difference between the test characteristic curve. These constants were then used to transform the Dutch discrimination (α) and location (β) parameters of the DIF items into new item parameters (α_{new} and β_{new}) on the US metric.

The mean T-score of respondents was calculated for the original and hybrid approach, to investigate the impact on T-scores on a group level. Furthermore, for each respondent the absolute difference between the original and hybrid approach was calculated, to investigate the impact on T-scores of individuals. To investigate the impact of DIF on CATs, it was assessed how often the DIF items were included in CATs, based on 4047 CATs for anxiety and 4293 CATs for depression from an ongoing study [32].

To provide reference values for the Dutch general population, T-scores on the complete item banks were calculated with the original US item parameters for the entire group of participants, as well as for age-range (18–34 years, 35–44 years, 45–54 years, 55–64 years, 65–74 years and $\geq$75 years) and gender subpopulations, in accordance with available subpopulation reference scores of the US population [56]. T-scores of the Dutch general population were compared to the US general population and age-range and gender subpopulation reference scores. T-score ranges that correspond to within normal limits, and to mild, moderate and severe symptoms [57] were computed using thresholds based on mean plus 0.5, 1 and 2 standard deviations. Subsequently, the percentage of participants that would fall within each category was calculated.

Results

A total of 1486 participants were invited, of which 1055 completed the PROMIS Anxiety and Depression item banks (response rate 71%). Because of suspicious response patterns (e.g. all responses in one category combined with short response times), 53 participants were excluded from the analysis. Sociodemographic characteristics of the remaining 1002 participants are presented in Table 1. Differences in sociodemographic characteristics between the study participants and the Dutch general population in 2013 were all less than 2.5%, except for ethnicity.

Monte Carlo simulations indicated that the type I error rate for DIF detection was well controlled, as the empirical thresholds for probability associated with the χ^2 statistic were all close to the nominal α (= 0.01) level, ranging from 0.009–0.01 for both item banks. This indicates that there is no need for establishing empirical thresholds through Monte Carlo simulations [51]. The McFadden's pseudo R^2 thresholds from the Monte Carlo simulations were all very small ($\leq$0.0004 for anxiety and $\leq$0.0003 for depression), and as this is an effect size measure, applying a threshold that is substantially less than what would be considered a small but

Table 1. Sociodemographic characteristics of participants and the Dutch general population.

Sociodemographic characteristic	Study participants* (n = 1002)	Dutch adult population 2013[a] (n = 13.3 million)
Age in years, mean ± SD (range)	49 ± 17 (18–100)	
18–39	34.3	34
40–64	44.4	44
≥65	21.3	22
Gender		
Male	47.9	49
Female	52.1	51
Educational level		
Low	32.0	32
Middle	39.9	40
High	28.0	28
Region of residence		
North	11.5	10
East	20.5	21
South	21.5	22
West	46.6	47
Ethnicity		
Native	79.6	80
1st and 2nd generation western immigrant	12.6	10
1st and 2nd generation non-western immigrant	7.8	10
Living situation		
Single	29.2	
Married/living together	60.0	
Relationship, not living together	4.0	
Living with parents	5.7	
Other	1.1	
Currently treated for psychological complaints		
Yes	10.1	
No	89.9	

* all results expressed as % unless otherwise noted.

SD: standard deviation;

[a] Based on data from statistics Netherlands (https://www.cbs.nl)

https://doi.org/10.1371/journal.pone.0273287.t001

meaningful effect (e.g. 0.02) would not be meaningful according to any standard [51]. Therefore, the nominal α level of 0.01 and the McFadden's pseudo R^2 value of 0.02 were maintained. Table 2 shows the results of the DIF analyses. Two items in the anxiety item bank, 'It scared me when I felt nervous' (EDANX03) and 'I felt worried' (EDANX30), showed uniform and non-uniform DIF, respectively. The item 'I felt worried' is present in the PROMIS anxiety 7a short form. The items are present in respectively 1 and 3% of the CAT-based assessments. In the depression item bank, two items showed uniform DIF: 'I felt worthless' (EDDEP04) and 'I felt unhappy' (EDDEP36). Both these items are present in the PROMIS depression 6a, 8a and 8b short forms. The item 'I felt worthless' is also present in the PROMIS depression 4a short form. The item 'I felt worthless' is present in 8% of the CAT-based assessments, whereas the item 'I felt unhappy' is present in all CAT-based assessments. For the item 'It scared me when I felt nervous', the threshold parameters for the Dutch population were mostly slightly lower than the thresholds for the US population, indicating that the Dutch population endorses

Table 2. McFadden's pseudo R^2 and IRT parameters for items displaying DIF.

Item bank	Item with DIF	DIF type	McFadden's pseudo R^2	Slope; and threshold parameters	Included in CAT[d]
Anxiety	EDANX03: It scared me when I felt nervous	Uniform	R^2_{12} = 0.021 R^2_{23} = 0.011	**NL**: 2.62; 0.15, 0.97, 2.02	1%
				US: 3.74; 0.59, 1.18, 1.95	
	EDANX30: I felt worried[a]	Non-uniform	R^2_{12} = 0.010 R^2_{23} = 0.033	NL: 2.16; -1.12, -0.10, 1.29, 2.64 US: 3.14; -0.57, 0.24, 1.22, 2.12	3%
Depression	EDDEP04: I felt worthless[b]	Uniform	R^2_{12} = 0.024 R^2_{23} = 0.013	**NL**: 2.93; -0.17, 0.58, 1.56, 2.61 US: 4.37; 0.29, 0.88, 1.61, 2.36	8%
	EDDEP36: I felt unhappy[c]	Uniform	R^2_{12} = 0.037 R^2_{23} = 0.001	NL: 4.21; -0.14, 0.61, 1.33, 2.20 **US**: 3.44; -0.64, 0.23, 1.20, 2.17	100%

The bold population had lower thresholds compared to the other population, indicating that this population endorses higher item response categories at the same level of the domain (anxiety, depression)

[a] present in the anxiety 7a short form

[b] present in the depression 4a, 6a, 8a and 8b short form

[c] present in the depression 6a, 8a and 8b short form

[d] Based on 4047 CAT-based assessments for anxiety and 4293 CAT-based assessments for depression

https://doi.org/10.1371/journal.pone.0273287.t002

higher response categories at the same level of anxiety. The same applied for the item 'I felt worthless'. For the item 'I felt unhappy', the threshold parameters for the Dutch population were slightly higher than the thresholds for the US population, indicating that the Dutch population endorses lower response categories at the same level of depression. Fig 1 illustrates the impact of DIF on respondents total scores. The plots on the left show the impact of DIF when all items are considered, whereas the plots on the right show the impact of DIF when only DIF items are considered. The plots show that DIF had a minimal impact on the total score when all items are administered in each item bank. S1 Fig shows the impact of DIF on item scores per group for the items displaying DIF.

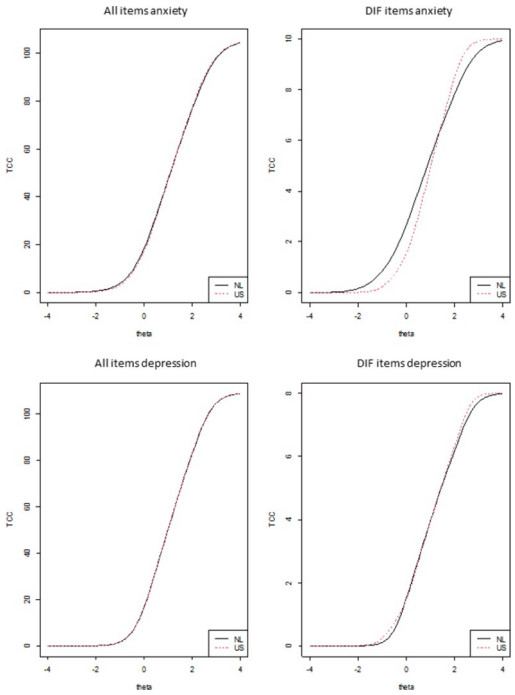

Fig 1. Total impact of DIF on the test characteristic curve (TCC) for anxiety and depression.

https://doi.org/10.1371/journal.pone.0273287.g001

Table 3 displays the impact of DIF on T-scores of item banks, and short forms wherein DIF items are present. On a population level, mean anxiety T-scores based on hybrid parameters were approximately 0.5 point lower than T-scores based on the original US parameters, both for the full item bank as the short form. Differences on a population level were even smaller for depression T-scores of item banks and most short forms. Only for the 4a short form, mean depression T-scores based on the hybrid parameters were 1 point lower than T-scores based on the original US parameters. On an individual level, absolute T-score differences between the original and hybrid approach for anxiety ranged from 0 to 1.7 for the full item bank and from 0 to 1.9 for the short form. Absolute T-score differences between the original and hybrid approach for the depression item banks and most short forms ranged from 0 to 1.2 for individuals. A maximum T-score difference between the two approaches of 2.6 was found for the depression short form 4a. S2 Fig shows the difference for each full item bank and short form in relation to the T-score of individuals. Notably, the largest differences were found for participants with T-scores on the lower end of the scale.

Dutch reference values for anxiety and depression and comparisons with the US general population, using the original US item parameters, are presented in Table 4. Differences between T-scores of the Dutch general population and T-scores of the US general population for anxiety and depression were small (difference of 0.1 and 0.4 for anxiety and depression, respectively). Differences between T-scores of the Dutch general population and T-scores of the US general population for age-range and gender subpopulations were also small (differences between 0.1 and 0.7 for anxiety and between 0.1 and 1.4 for depression). T-scores of the Dutch general population and US general population showed similar patterns, with males scoring lower (i.e. less anxious or depressed) than females and lower scores for older age groups.

Using 0.5, 1 and 2 standard deviations, thresholds for mild, moderate and severe anxiety were set to 55, 60 and 70 respectively. The same thresholds applied for depression (Fig 2). When these thresholds were applied to anxiety T-scores of participants, 70% fell within normal limits (i.e. $\leq$55), 14% had mild symptoms (i.e. 56–60), 15% had moderate symptoms (i.e. 61–70) and 1% had severe symptoms (i.e. >70). For depression, 71% fell within normal limits (i.e. $\leq$55), 15% had mild symptoms (i.e. 56–60), 13% had moderate symptoms (i.e. 61–70) and 1% had severe symptoms (i.e. >70).

Table 3. PROMIS anxiety and depression T-scores[a] based on different sets of item parameters for different versions of the instruments.

Version	Mean population T-score (SD) original approach[b]	Mean population T-score (SD) hybrid approach[c]	Mean absolute T-score difference (SD), [range]
Anxiety			
Full item bank	49.9 (10.1)	49.5 (10.3)	0.40 (0.29) [0.00–1.67]
Short form 7a	50.3 (9.2)	49.8 (9.6)	0.59 (0.48) [0.00–1.88]
Depression			
Full item bank	49.6 (10.0)	49.7 (9.9)	0.15 (0.09) [0.00–0.73]
Short form 4a	50.9 (8.5)	49.9 (8.7)	1.02 (0.52) [0.00–2.57]
Short form 6a	49.7 (9.3)	49.8 (9.0)	0.52 (0.29) [0.00–1.24]
Short form 8a	50.3 (9.2)	50.4 (8.9)	0.42 (0.25) [0.01–1.10]
Short form 8b	50.2 (9.3)	50.2 (9.1)	0.38 (0.22) [0.00–1.15]

SD: standard deviation

[a] T-scores, higher scores represent more anxiety/depression

[b] All items have the US item parameters

[c] Non-DIF items have the US item parameters, DIF items have the Dutch item parameters rescaled to the US metric

https://doi.org/10.1371/journal.pone.0273287.t003

Table 4. PROMIS anxiety and depression Dutch reference values[a] by age and gender and comparisons with the US general population [58].

		Anxiety			Depression		
	N Dutch population (%)	N US population (%)	Dutch mean T-score (SD)	US mean T-score (SD)	N US population (%)	Dutch mean T-score (SD)	US mean T-score (SD)
Total	1002 (100)	2724 (100)	49.9 (10.1)	50.0 (10.0)	2160 (100)	49.6 (10.0)	50.0 (10.0)
Gender							
Male	480 (48)	1069 (39)	49.0 (10.0)	48.6 (9.5)	890 (41)	48.8 (10.1)	48.7 (9.7)
Female	522 (52)	1654 (61)	50.6 (10.1)	50.9 (10.2)	1269 (59)	50.4 (9.9)	50.9 (10.1)
Age in years							
18–34	253 (25)	659 (24)	51.8 (9.9)	52.4 (10.7)	496 (23)	52.0 (9.3)	52.3 (10.9)
35–44	147 (15)	496 (18)	51.4 (10.9)	50.9 (11.1)	366 (17)	50.5 (10.8)	50.6 (10.9)
45–54	173 (17)	417 (15)	50.0 (10.9)	50.1 (9.5)	359 (17)	50.0 (11.0)	50.8 (10.0)
55–64	216 (22)	442 (16)	48.9 (9.4)	49.3 (9.5)	373 (17)	48.8 (9.8)	49.5 (9.7)
65–74	191 (19)	365 (13)	47.5 (9.1)	48.1 (8.8)	290 (13)	47.0 (8.9)	48.4 (8.8)
75+	22 (2)	345 (13)	46.2 (9.4)	46.9 (7.9)	276 (13)	46.0 (9.6)	46.5 (7.2)

SD: standard deviation

[a] T-scores, higher scores represent more anxiety/depression; T-scores were calculated based on the original US item parameters

https://doi.org/10.1371/journal.pone.0273287.t004

Discussion

This study assessed DIF for language between the Netherlands and the US for the PROMIS Anxiety and Depression item banks, and presented Dutch reference values for the general population and relevant subpopulations. We found some items with DIF, but the impact of DIF on population level T-scores was small, both for full item banks as for short forms. This supports the applicability of the US scoring algorithm in the Netherlands and strengthens the cross-cultural validity of the Dutch-Flemish PROMIS Anxiety and Depression item banks. It enables the comparison of scores between Dutch and US populations. The established Dutch reference values can be used to interpret symptoms of anxiety and depression in research and clinical practice.

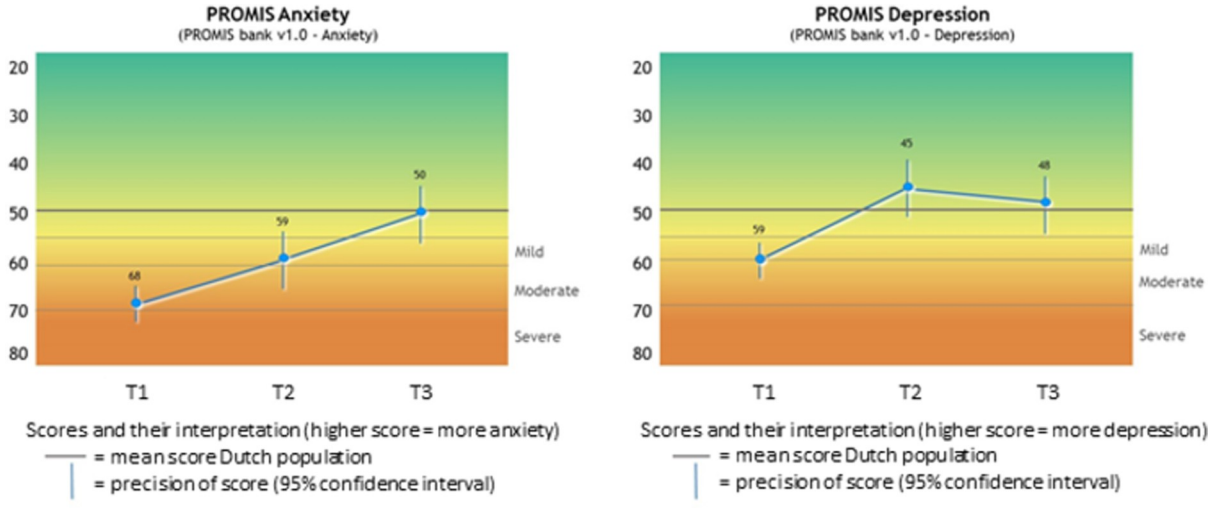

Fig 2. Visual feedback [59] of PROMIS Anxiety and Depression scores, based on Dutch mean T-scores and Dutch thresholds for mild, moderate and severe symptoms. The blue lines represent imaginary data showing the course of symptoms over three consecutive assessments (T1, T2 and T3).

https://doi.org/10.1371/journal.pone.0273287.g002

We only found a limited number of items with DIF, which had negligible impact on total scores when all items in the item banks were administered. However, the impact of DIF might be more in short forms wherein DIF items were present, because only a small number of items are administered. On a population level, the impact of DIF on T-scores was small for all short forms. Because DIF in the two depression items had opposite direction, the effect of DIF might have been canceled out in the short forms in which both items were present. On individual level the impact was larger, especially for the depression short form 4a, with a maximum difference of 2.6 points. This is close to the amount of 3 points that is generally considered minimally important [60–63], and therefore this short form might not be the best option to assess symptoms of depression in individuals. Most DIF items were not frequently administered in CAT-based assessments, but the item 'I felt unhappy' was present in all CATs [32]. In a future study, it might be interesting to explore the impact of DIF on CAT T-scores and also to explore whether omitting DIF items from the item bank could result in equally precise scores and similar amount of items administered.

DIF for language could be caused by a lack of translational equivalence [64]. In the development of PROMIS measures generally a translatability review is performed, but only for Spanish. In a translatability review, the original measure is reviewed to determine its suitability for future translations. A translatability review is best conducted as early as possible during the development of a new measure, preferably before quantitative testing, as changes to the measure can still be made at this point. To make new PROMIS measures more applicable for translations to other languages, which is increasingly occurring, a broader translatability review might be a useful additional step in the developmental process of PROMIS measures. During the Dutch-Flemish translation process of the PROMIS Anxiety and Depression item banks, no particular difficulties were experienced with translating the items showing DIF for language [42]. The items 'I felt worried' and 'I felt unhappy' also showed DIF for language in a study comparing the Brazilian to the US version [41], but studies in Germany and Spain found different DIF items [65, 66]. Although a translatability review might reduce translation difficulties, it does not replace the evaluation of DIF for language, as DIF can also occur due to cultural differences [67]. Therefore DIF studies are recommended after every translation [52].

The negligible impact of DIF made it possible to compare item bank scores of the Dutch general population to the US general population. T-scores of the Dutch general population were similar to scores of the US general population, both for the total population (difference of 0.1 for anxiety and 0.4 for depression) and for age-range and gender subpopulations. Unfortunately, it is not clear yet what a minimal important difference is in scores between groups for anxiety and depression [68], although most studies suggest a within-person change of at least 3 points to be meaningful [60–63]. Thus, we think it is safe to conclude that T-scores of the Dutch general population were similar to scores of the US general population. Because of similarity in scores and standard deviations, thresholds for mild, moderate and severe symptoms of anxiety and depression were identical to the thresholds based on the US data [57, 63].

The inclusion of anxiety and depression outcomes in many ICHOM Standard Sets shows that measuring anxiety and depression is relevant for many patient groups and persons without diseases, and not only those with mental disorders [6, 7]. In the Standard Set for Overall Adult Health, it is advocated to measure anxiety and depression via the PROMIS Scale v1.2 – Global Health [69], resulting in a global mental health score [70], for which Dutch reference values recently have been published [48]. In the other 16 Standard Sets that include anxiety and depression, a range of PROMs is advocated, including disease-specific PROMs, cancer-specific PROMs, anxiety/depression-specific PROMs and generic PROMs [6, 7]. A more universal and standardized approach to measuring anxiety and depression will facilitate outcome measurement in clinical practice and comparisons of scores across patient groups [14, 15].

PROMIS anxiety and depression instruments offer opportunities here, and the results of this study expands their utility.

PROMIS anxiety and depression instruments have several advantages over current legacy instruments for anxiety and depression. First, PROMIS instruments are applicable across the general population and various patient groups, as well as those patients with multimorbidity, rare diseases, or without a definite diagnoses [17, 18, 20]. This enables the comparison of patient groups, benchmarking and improving the quality of care. Second, PROMIS Anxiety and Depression item banks can be used as CAT, which reduces the response burden while high measurement precision is maintained, and as such is valuable in clinical practice [23]. Currently a limited number of countries, including the Netherlands, have access to technical solutions for CAT applications, but this is expected to expand rapidly in the near future. Third, several crosswalk studies have linked scores of legacy instruments to PROMIS anxiety and depression instruments [71–75], which facilitates the uptake of PROMIS instruments and the interpretation of scores, even when legacy instruments have been used in the past. Last, PROMIS is a sustainable state-of-the-art measurement system that is actively maintained by the PROMIS Health Organization, in order to facilitate the widespread use and adoption of PROMIS in research and clinical practice.

A strength of the present study is that we not only assessed the impact of DIF when all items were considered, but applied Stocking-Lord constants to investigate the impact of DIF on T-scores of full item banks and short forms. Moreover, the large sample size made sure that the Dutch reference values have been estimated reliably. However, some subgroups (especially adults ages 75 years and older) were relatively small, which can be considered a limitation. Second, although our sample was broadly representative for the Dutch general population on some important characteristics, we cannot be certain that this is also the case for other important characteristics, such as income level and employment status. One could argue that persons who have the time to participate in an internet panel and complete item banks, might more often be persons without full-time employment, which might in turn be caused by physical or mental problems. The non-probabilistic selection procedure might have had an impact on the general population reference scores presented in this article.

Conclusions

The limited number of items with DIF in PROMIS Anxiety and Depression item banks, having small impact on population T-scores, supports the applicability of the US scoring algorithm and enables the comparison of scores of the Dutch and US population. The Dutch general population had a T-score of 49.9 for anxiety and 49.6 for depression, similar to the T-scores of 50 of the US general population. The Dutch reference values reported in this study provide an important tool for healthcare professionals and researchers to evaluate and interpret symptoms of anxiety and depression. The presented reference values for subpopulations allow a more tailored and relevant interpretation and understanding of symptoms of anxiety and depression. Incorporating the Dutch reference values and thresholds in the feedback patients and healthcare professionals receive regarding their mental health status as assessed with PROMIS anxiety and depression instruments, will facilitate interpretation of scores by patients and healthcare professionals. The availability of Dutch reference values may stimulate the uptake of PROMIS instruments for anxiety and depression, and contribute to standardized measurements of anxiety and depression.

Supporting information

S1 Fig. Category response curves for items displaying DIF.
(TIF)

S2 Fig. Relation between T-scores and differences in T-scores original vs. hybrid approach.
(TIF)

S1 Data. Anxiety data of respondents.
(POR)

S2 Data. Depression data of respondents.
(POR)

Author Contributions

Conceptualization: Edwin de Beurs, Leo D. Roorda, Caroline B. Terwee.

Data curation: Gerard Flens.

Formal analysis: Ellen B. M. Elsman.

Investigation: Ellen B. M. Elsman, Gerard Flens, Caroline B. Terwee.

Methodology: Caroline B. Terwee.

Resources: Caroline B. Terwee.

Supervision: Caroline B. Terwee.

Validation: Gerard Flens.

Visualization: Ellen B. M. Elsman.

Writing – original draft: Ellen B. M. Elsman.

Writing – review & editing: Gerard Flens, Edwin de Beurs, Leo D. Roorda, Caroline B. Terwee.

References

1. Lloyd C, Dyer P, Barnett A. Prevalence of symptoms of depression and anxiety in a diabetes clinic population. Diabetic medicine. 2000; 17(3):198–202. https://doi.org/10.1046/j.1464-5491.2000.00260.x PMID: 10784223
2. Singer S, Das-Munshi J, Brähler E. Prevalence of mental health conditions in cancer patients in acute care—a meta-analysis. Annals of Oncology. 2010; 21(5):925–30. https://doi.org/10.1093/annonc/mdp515 PMID: 19887467
3. Hare DL, Toukhsati SR, Johansson P, Jaarsma T. Depression and cardiovascular disease: a clinical review. European heart journal. 2014; 35(21):1365–72. https://doi.org/10.1093/eurheartj/eht462 PMID: 24282187
4. Frank JD. Psychotherapy: The restoration of morale. American Journal of Psychiatry. 1974; 131 (3):271–4. https://doi.org/10.1176/ajp.131.3.271 PMID: 4812687
5. Clarke DM, Kissane DW. Demoralization: its phenomenology and importance. Australian & New Zealand Journal of Psychiatry. 2002; 36(6):733–42. https://doi.org/10.1046/j.1440-1614.2002.01086.x PMID: 12406115
6. ICHOM. International Consortium for Health Outcomes Measurement (ICHOM) 2021. Available from: www.ichom.org.
7. Terwee CB, Zuidgeest M, Vonkeman HE, Cella D, Haverman L, Roorda LD. Common patient-reported outcomes across ICHOM Standard Sets: the potential contribution of PROMIS®. BMC Medical Informatics and Decision Making. 2021; 21(1):1–13.
8. Nezu AM, Ronan GF, Meadows EA, McClure KS. Practitioner's guide to empirically-based measures of depression: Springer Science & Business Media; 2000.
9. Roemer L. Measures for anxiety and related constructs. Practitioner's guide to empirically based measures of anxiety: Springer; 2002. p. 49–83.

10. Antony MM, Stein MB. Oxford handbook of anxiety and related disorders: Oxford University Press; 2008.
11. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. Journal of the National Cancer Institute. 2009; 101(21):1464–88. https://doi.org/10.1093/jnci/djp336 PMID: 19826136
12. McHugh RK, Rasmussen JL, Otto MW. Comprehension of self-report evidence-based measures of anxiety. Depression and Anxiety. 2011; 28(7):607–14. https://doi.org/10.1002/da.20827 PMID: 21618668
13. Nelson CJ, Cho C, Berk AR, Holland J, Roth AJ. Are gold standard depression measures appropriate for use in geriatric cancer patients? A systematic evaluation of self-report depression instruments used with geriatric, cancer, and geriatric cancer samples. Journal of Clinical Oncology. 2010; 28(2):348. https://doi.org/10.1200/JCO.2009.23.0201 PMID: 19996030
14. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ. 2019;364. https://doi.org/10.1136/bmj.k5267 PMID: 30679170
15. Jim HS, Hoogland AI, Brownstein NC, Barata A, Dicker AP, Knoop H, et al. Innovations in research and clinical care using patient-generated health data. CA: a cancer journal for clinicians. 2020; 70(3):182–99. https://doi.org/10.3322/caac.21608 PMID: 32311776
16. Eton DT, Beebe TJ, Hagen PT, Halyard MY, Montori VM, Naessens JM, et al. Harmonizing and consolidating the measurement of patient-reported information at health care institutions: a position statement of the Mayo Clinic. Patient Related Outcome Measures. 2014; 5:7. https://doi.org/10.2147/PROM.S55069 PMID: 24550683
17. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. Journal of clinical epidemiology. 2010; 63(11):1179–94. https://doi.org/10.1016/j.jclinepi.2010.04.011 PMID: 20685078
18. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Medical care. 2007; 45(5 Suppl 1):S3. https://doi.org/10.1097/01.mlr.0000258615.42478.55 PMID: 17443116
19. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Maihoefer CC, Lawrence SM. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS®) in a three-month observational study. Journal of psychiatric research. 2014; 56:112–9. https://doi.org/10.1016/j.jpsychires.2014.05.010 PMID: 24931848
20. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. Journal of clinical epidemiology. 2016; 73:119–27. https://doi.org/10.1016/j.jclinepi.2015.08.036 PMID: 26931289
21. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment. 2011; 18(3):263–83. https://doi.org/10.1177/1073191111411667 PMID: 21697139
22. Embretson SE, Reise SP. Item response theory: Psychology Press; 2013.
23. Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Quality of Life Research. 2007; 16(1):133–41.
24. HealthMeasures. PROMIS Anxiety Scoring Manual [cited 2020]. Available from: https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Anxiety_Scoring_Manual.pdf.
25. HealthMeasures. PROMIS Depression Scoring Manual [cited 2020]. Available from: https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Depression_Scoring_Manual.pdf.
26. HealthMeasures. How to select a HealthMeasure 2020 [cited 2020 December]. Available from: https://www.healthmeasures.net/applications-of-healthmeasures/guidance/selecting-a-healthmeasure.
27. Purvis TE, Neuman BJ, Riley LH, Skolasky RL. Comparison of PROMIS Anxiety and Depression, PHQ-8, and GAD-7 to screen for anxiety and depression among patients presenting for spine surgery. Journal of Neurosurgery: Spine. 2019; 30(4):524–31.
28. Sunderland M, Batterham P, Calear A, Carragher N. Validity of the PROMIS depression and anxiety common metrics in an online sample of Australian adults. Quality of Life Research. 2018; 27(9):2453–8. https://doi.org/10.1007/s11136-018-1905-5 PMID: 29872956
29. Clover K, Lambert SD, Oldmeadow C, Britton B, King MT, Mitchell AJ, et al. PROMIS depression measures perform similarly to legacy measures relative to a structured diagnostic interview for depression in cancer patients. Quality of Life Research. 2018; 27(5):1357–67. https://doi.org/10.1007/s11136-018-1803-x PMID: 29423755

30. Amtmann D, Kim J, Chung H, Bamer AM, Askew RL, Wu S, et al. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. Rehabilitation psychology. 2014; 59(2):220. https://doi.org/10.1037/a0035919 PMID: 24661030

31. Freedland KE, Steinmeyer BC, Carney RM, Rubin EH, Rich MW. Use of the PROMIS® Depression scale and the Beck Depression Inventory in patients with heart failure. Health Psychology. 2019; 38 (5):369. https://doi.org/10.1037/hea0000682 PMID: 31045419

32. Flens G, Terwee CB, Smits N, Williams G, Spinhoven P, Roorda LD, et al. Construct validity, responsiveness, and utility of change indicators of the Dutch-Flemish PROMIS item banks for depression and anxiety administered as computerized adaptive test (CAT): A comparison with the Brief Symptom Inventory (BSI). Psychological Assessment. 2021.

33. Lizzio VA, Blanchett J, Borowsky P, Meldau JE, Verma NN, Muh S, et al. Feasibility of PROMIS CAT administration in the ambulatory sports medicine clinic with respect to cost and patient compliance: a single-surgeon experience. Orthopaedic journal of sports medicine. 2019; 7(1):2325967118821875. https://doi.org/10.1177/2325967118821875 PMID: 30733973

34. Beleckas CM, Prather H, Guattery J, Wright M, Kelly M, Calfee RP. Anxiety in the orthopedic patient: using PROMIS to assess mental health. Quality of Life Research. 2018; 27(9):2275–82. https://doi.org/10.1007/s11136-018-1867-7 PMID: 29740783

35. Papuga MO, Dasilva C, McIntyre A, Mitten D, Kates S, Baumhauer J. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. Health Systems. 2018; 7(1):1–12. https://doi.org/10.1057/s41306-016-0016-1 PMID: 31214335

36. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer. 2015; 121(6):927–34. https://doi.org/10.1002/cncr.29104 PMID: 25376427

37. Scholle SH, Morton S, Homco J, Rodriguez K, Anderson D, Hahn E, et al. Implementation of the PROMIS-29 in routine care for people with diabetes: challenges and opportunities. The Journal of ambulatory care management. 2018; 41(4):274–87. https://doi.org/10.1097/JAC.0000000000000248 PMID: 29923844

38. HealthMeasures. Available translations 2020 [cited 2020 December]. Available from: https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations.

39. Vilagut G, Forero C, Adroher N, Olariu E, Cella D, Alonso J, et al. Testing the PROMIS® Depression measures for monitoring depression in a clinical sample outside the US. Journal of psychiatric research. 2015; 68:140–50. https://doi.org/10.1016/j.jpsychires.2015.06.009 PMID: 26228413

40. Jakob T, Nagl M, Gramm L, Heyduck K, Farin E, Glattacker M. Psychometric properties of a German translation of the PROMIS® depression item bank. Evaluation & the health professions. 2017; 40 (1):106–20.

41. de Castro NFC, Pinto RdMC, da Silva Mendonça TM, da Silva CHM. Psychometric validation of PROMIS® Anxiety and Depression Item Banks for the Brazilian population. Quality of Life Research. 2020; 29(1):201–11. https://doi.org/10.1007/s11136-019-02319-1 PMID: 31598816

42. Terwee C, Roorda L, De Vet H, Dekker J, Westhovens R, Van Leeuwen J, et al. Dutch–Flemish translation of 17 item banks from the patient-reported outcomes measurement information system (PROMIS). Quality of Life Research. 2014; 23(6):1733–41. https://doi.org/10.1007/s11136-013-0611-6 PMID: 24402179

43. Flens G, Smits N, Terwee CB, Dekker J, Huijbrechts I, de Beurs E. Development of a computer adaptive test for depression based on the Dutch-Flemish version of the PROMIS item bank. Evaluation & the health professions. 2017; 40(1):79–105. https://doi.org/10.1177/0163278716684168 PMID: 28705028

44. Flens G, Smits N, Terwee CB, Dekker J, Huijbrechts I, Spinhoven P, et al. Development of a computerized adaptive test for anxiety based on the Dutch–Flemish version of the PROMIS item bank. Assessment. 2019; 26(7):1362–74. https://doi.org/10.1177/1073191117746742 PMID: 29231048

45. Flens G, Smits N, Terwee CB, Pijck L, Spinhoven P, de Beurs E. Practical Significance of Longitudinal Measurement Invariance Violations in the Dutch–Flemish PROMIS Item Banks for Depression and Anxiety: An Illustration With Ordered-Categorical Data. Assessment. 2021; 28(1):277–94. https://doi.org/10.1177/1073191119880967 PMID: 31625411

46. van Bebber J, Flens G, Wigman JT, de Beurs E, Sytema S, Wunderink L, et al. Application of the Patient-Reported Outcomes Measurement Information System (PROMIS) item parameters for Anxiety and Depression in the Netherlands. International journal of methods in psychiatric research. 2018; 27 (4):e1744. https://doi.org/10.1002/mpr.1744 PMID: 30246495

47. Fischer F, Gibbons C, Coste J, Valderas JM, Rose M, Leplège A. Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. Quality of Life Research. 2018; 27(4):999–1014. https://doi.org/10.1007/s11136-018-1785-8 PMID: 29350345

48. Elsman EB, Roorda LD, Crins MH, Boers M, Terwee CB. Dutch reference values for the Patient-Reported Outcomes Measurement Information System Scale v1. 2-Global Health (PROMIS-GH). Journal of Patient-Reported Outcomes. 2021; 5(1):1–9.

49. Cella D. PROMIS 1 Wave 1. Harvard Dataverse; 2015.

50. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical care. 2007:S22–S31. https://doi.org/10.1097/01.mlr.0000250483.85507.04 PMID: 17443115

51. Choi SW, Gibbons LE, Crane PK. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. Journal of statistical software. 2011; 39(8):1. https://doi.org/10.18637/jss.v039.i08 PMID: 21572908

52. HealthMeasures. Minimum requirements for the release of PROMIS instruments after translation and recommendations for further psychometirc evaluation. 2014.

53. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. Journal of statistical Software. 2012; 48(1):1–29.

54. Stocking ML, Lord FM. Developing a common metric in item response theory. Applied psychological measurement. 1983; 7(2):201–10.

55. Kolen MJ, Brennan RL. Test equating, scaling, and linking: Methods and practices: Springer Science & Business Media; 2014.

56. HealthMeasures. PROMIS reference populations 2021 [cited 2021 March]. Available from: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/reference-populations.

57. HealthMeasures. PROMIS Score Cut-Points [cited 2020]. Available from: http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points.

58. HealthMeasures. Gender and Age Range Sub-norms for Adult PROMIS Measures Centered on the US General Census 2000 [cited 2020]. Available from: http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/reference-populations.

59. van Muilekom MM, Luijten MA, van Oers HA, Terwee CB, van Litsenburg RR, Roorda LD, et al. From statistics to clinics: the visual feedback of PROMIS® CATs. Journal of Patient-Reported Outcomes. 2021; 5(1):1–14.

60. Swanholm E, McDonald W, Makris U, Noe C, Gatchel R. Estimates of Minimally Important Differences (MID s) for Two Patient-Reported Outcomes Measurement Information System (PROMIS) Computer-Adaptive Tests in Chronic Pain Patients. Journal of Applied Biobehavioral Research. 2014; 19(4):217–32.

61. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. Journal of clinical epidemiology. 2011; 64(5):507–16. https://doi.org/10.1016/j.jclinepi.2010.11.018 PMID: 21447427

62. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and minimally important differences for 4 patient-reported outcomes measurement information system short forms: physical function, pain interference, depression, and anxiety in knee osteoarthritis. The Journal of Pain. 2017; 18(9):1096–110. https://doi.org/10.1016/j.jpain.2017.05.001 PMID: 28501708

63. Kroenke K, Stump TE, Chen CX, Kean J, Bair MJ, Damush TM, et al. Minimally important differences and severity thresholds are estimated for the PROMIS depression scales from three randomized clinical trials. Journal of affective disorders. 2020; 266:100–8. https://doi.org/10.1016/j.jad.2020.01.101 PMID: 32056864

64. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. Health and quality of life outcomes. 2010; 8(1):1–9. https://doi.org/10.1186/1477-7525-8-81 PMID: 20684767

65. Vilagut G, Forero CG, Castro-Rodriguez JI, Olariu E, Barbaglia G, Astals M, et al. Measurement equivalence of PROMIS depression in Spain and the United States. Psychological assessment. 2019; 31(2):248. https://doi.org/10.1037/pas0000665 PMID: 30382716

66. Fischer HF, Wahl I, Nolte S, Liegl G, Brähler E, Löwe B, et al. Language-related differential item functioning between English and German PROMIS Depression items is negligible. International journal of methods in psychiatric research. 2017; 26(4):e1530. https://doi.org/10.1002/mpr.1530 PMID: 27747969

67. Acquadro C, Patrick DL, Eremenco S, Martin ML, Kuliś D, Correia H, et al. Emerging good practices for translatability assessment (TA) of patient-reported outcome (PRO) measures. Journal of patient-reported outcomes. 2018; 2(1):1–11.

68. HealthMeasures. Meaningful change for PROMIS [cited 2020]. Available from: http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/meaningful-change.

69. ICHOM. Overall Adult Health 2021 [cited 2021 February]. Available from: https://www.ichom.org/portfolio/overall-adult-health/.

70. HealthMeasures. PROMIS Global Health Scoring Manual [cited 2020]. Available from: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Global_Scoring_Manual.pdf.

71. Victorson D, Schalet BD, Kundu S, Helfand BT, Novakovic K, Penedo F, et al. Establishing a common metric for self-reported anxiety in patients with prostate cancer: Linking the Memorial Anxiety Scale for Prostate Cancer with PROMIS Anxiety. Cancer. 2019; 125(18):3249–58. https://doi.org/10.1002/cncr.32189 PMID: 31090933

72. Kaat AJ, Newcomb ME, Ryan DT, Mustanski B. Expanding a common metric for depression reporting: linking two scales to PROMIS® depression. Quality of Life Research. 2017; 26(5):1119–28. https://doi.org/10.1007/s11136-016-1450-z PMID: 27815821

73. Kim J, Chung H, Askew RL, Park R, Jones SM, Cook KF, et al. Translating CESD-20 and PHQ-9 scores to PROMIS depression. Assessment. 2017; 24(3):300–7. https://doi.org/10.1177/1073191115607042 PMID: 26423348

74. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychological assessment. 2014; 26(2):513. https://doi.org/10.1037/a0035768 PMID: 24548149

75. Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. Journal of anxiety disorders. 2014; 28(1):88–96. https://doi.org/10.1016/j.janxdis.2013.11.006 PMID: 24508596