

Promising PROMIS

Psychometric properties of the
Dutch-Flemish PROMIS Pain
Behavior, Pain Interference and
Physical Function item banks

Martine H.P. Crins

Promising PROMIS
Psychometric properties of
the Dutch-Flemish PROMIS Pain Behavior,
Pain Interference and
Physical Function item banks

Contents

- 7 Chapter 1** General introduction
- 19 Chapter 2** Calibration of the Dutch-Flemish PROMIS Pain Behavior Item Bank in Patients with Chronic Pain
European Journal of Pain 2016;20(2):284-96
- 43 Chapter 3** Calibration and validation of the Dutch-Flemish PROMIS Pain Interference Item Bank in Patients with Chronic Pain
Plos One 2015;10(7):1-18
- 67 Chapter 4** The Dutch-Flemish PROMIS Physical Function Item Bank Exhibited Strong Psychometric Properties in Patients with Chronic Pain
Journal of Clinical Epidemiology 2017;87:47-58
- 95 Chapter 5** First Validation of the Full PROMIS Pain Interference and Pain Behavior Item Banks in Patients with Rheumatoid Arthritis
Arthritis Care Res 2019;Sep28
- 117 Chapter 6** Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy
PloS One 2018;13(2):1-14
- 137 Chapter 7** Differential item functioning of the PROMIS Physical Function, Pain Interference and Pain Behavior item banks across patients with different musculoskeletal disorders and persons from the general population
Qual Life Res 2019;28(5):1231-1243
- 157 Chapter 8** International application of PROMIS® Computerized Adaptive Tests: US versus country-specific item parameters
Submitted
- 179 Chapter 9** General discussion
- 197 Chapter 10**
- 199 Summary
- 203 Samenvatting
- 207 Dankwoord
- 213 About the author
- 215 List of publications
- 217 PhD portfolio

Summary

Measurement of health as perceived by patients has become more and more important within healthcare. Patient-Reported Outcomes (PROs) are outcomes that reflect the self-perceived status of the patient's health without interpretation by a clinician or anyone else. Examples of PROs are constructs (or domains) like physical function, pain, fatigue, anxiety, and depression. Patient-Reported Outcome Measures (PROMs) are questionnaires used to measure PROs.

Many traditional PROMs are not without shortcomings. They are burdensome to patients, since patients have to answer all questions while not all questions are always relevant for them, have large measurement error and the scores are often difficult to interpret and cannot be compared between the many existing PROMs.

PROMs based on Item Response Theory (IRT) have the potential to overcome most of these shortcomings of traditional PROMs. Measures developed using IRT methods consist of item banks, a set of items (questions) measuring one construct (domain). Items can be selected from an item bank to be administered as short forms (fixed length subsets of items out of the item bank) or highly efficient computer adaptive test (CAT). A CAT is a computer-administered measure in which, after the first item, successive items are selected by a computer algorithm based on persons' responses to previous items.

The Patient-Reported Outcomes Measurement Information System (PROMIS®) is the largest currently available system of IRT-based item banks, short forms and CATs for measuring PROs. Advantages of PROMIS are that it has been developed using a comprehensive framework, that PROMIS instruments are generic instead of disease specific, that PROMIS instruments already showed to have better content validity and smaller measurement error than traditional disease specific PROMs, and that PROMIS scores are easy to interpret and can be used to compare (sub)groups within and between populations. PROMIS has been developed in the United States (US) and has already been translated into several languages, including Dutch-Flemish.

Chapter 1 provides a general introduction to the research of this thesis and describes which steps are required to apply PROMIS in the Netherlands and Flanders. The scope and outline of the thesis are given and the main aim of this thesis is to perform a first exploration whether PROMIS can be applied to measure health in Dutch-Flemish populations.

Chapters 2-6 present the validation of the Dutch-Flemish PROMIS Pain Behavior (v1.1), Pain Interference (v1.1) and Physical Function (v1.2) item banks in multiple Dutch and Flemish patient populations. All three PROMIS item banks show good psychometric properties in Dutch patients with chronic pain (**Chapters 2, 3 and 4**, respectively). Furthermore, the Dutch-Flemish PROMIS Pain Behavior and Pain Interference item banks show good psychometric properties in Dutch and Flemish patients with

rheumatoid arthritis as well (**Chapter 5**). The Dutch-Flemish PROMIS Physical Function item bank shows also good psychometric properties in Dutch patients receiving physical therapy (**Chapter 6**). All three item banks meet the IRT-model assumptions and show good IRT-model fit. The item banks show good cross-cultural validity (no Differential Item Functioning (DIF) for language; Dutch-Flemish vs. English and Dutch vs. Flemish), absence of other measurement variance (no DIF for age, gender, administration mode and/or disease activity), good construct validity, high reliability in the range of the patient samples, and absence of floor and ceiling effects.

In addition to the studies addressing the psychometric properties and in order to investigate the validity of comparisons across populations, the item parameters of the Dutch-Flemish PROMIS Pain Behavior, Pain Interference and Physical Function item banks are compared across several clinical samples with different musculoskeletal disorders and a general population sample (**Chapter 7**). The results show some items with DIF between some of these samples, but the impact of DIF on PROMIS T-scores is negligible. This study provides evidence for the universal applicability of the three item banks. The item banks and their scores can be used for valid comparisons across clinical samples with different musculoskeletal disorders and the general population.

Although it is the current PROMIS recommendation to use a single set of item parameters (current recommendation are the US item parameters) in the scoring of PROMIS instruments and in the PROMIS CAT algorithms across the world, there might be reasons to consider country-specific parameters in non-English speaking countries, like in the Netherlands or Flanders. Non-English PROMIS researchers may argue that the most accurate way to measure a certain population is to use item parameters obtained in a large sample of the population at issue. However, it is important from an international point of view, to obtain scores that are internationally comparable. In **Chapter 8**, a simulation study illustrates the consequences of choosing different sets of item parameters in PROMIS CATs in non-English speaking countries. Four options were examined that reflect different decisions around the retention or rescaling (to the US PROMIS metric) of country-specific IRT parameters. Based on the study results, the conclusion is that no evidence was found for changing the current PROMIS recommendations for group level comparisons. CAT T-scores obtained with US item parameters for DIF-free items and country-specific item parameters, rescaled to the US PROMIS metric, for DIF items, or CAT T-scores obtained with country-specific item parameters for all items, rescaled to the US PROMIS metric, are comparable to CAT T-scores obtained using the US item parameters on group level. However, the choice of item parameters can be consequential for individual patient scores. Using US or country-specific item parameters both have advantages and disadvantages. The PROMIS Health Organization (PHO) should decide upon preferred and allowed strategies for choosing item parameter sets outside of the US, to ensure consistency of reporting of PROMIS T-scores across populations and countries.

Chapter 9 presents the general discussion of this thesis, shows what this thesis brought forth and provides some recommendations for future research. Issues relevant

for the implementation of PROMIS in the Netherlands and Flanders are addressed, and several examples of implementations of PROMIS CATs in daily clinical practice are given.

This thesis is an important step for the use of PROMIS in the Netherlands and Flanders. It is part of a large national PROMIS initiative. The current thesis contributes to the evidence that PROMIS consists of a set of standardized high quality PROMs. It shows good psychometric quality of PROMIS item banks in multiple patient samples. The three investigated Dutch-Flemish PROMIS item banks are appropriate and ready for the use in the Netherlands and Flanders according to the PROMIS requirements for the release of PROMIS instruments in languages other than English and in countries other than the US. PROMIS can be applied well in the Netherlands and Flanders. Worldwide, great efforts are made to implement PROMIS in daily clinical practice and in research. Because of this, health can be measured efficiently and patient-friendly.

Samenvatting

Het meten van gezondheid zoals ervaren door patiënten is in de gezondheidszorg steeds belangrijker geworden. Patiënt-gerapporteerde uitkomsten (PROs) zijn uitkomsten die de zelf-ervaren gezondheid weergeven, zonder interpretatie door een arts of iemand anders. Voorbeelden van PROs zijn constructen (of domeinen) zoals fysiek functioneren, pijn, vermoeidheid, angst en depressie. Patiënt-gerapporteerde uitkomstmaten (PROMs) zijn vragenlijsten die worden gebruikt om PROs te meten.

Veel traditionele PROMs hebben tekortkomingen. Ze zijn erg belastend voor patiënten omdat patiënten alle vragen moeten beantwoorden, terwijl niet alle vragen altijd relevant zijn. Daarnaast hebben veel traditionele PROMs een grote meetfout, zijn de scores vaak moeilijk te interpreteren en kunnen de scores niet worden vergeleken tussen de verschillende PROMs.

PROMs gebaseerd op Item Response Theory (IRT) hebben de potentie om de meeste tekortkomingen van traditionele PROMs te overwinnen. Meetinstrumenten ontwikkeld met behulp van IRT-methoden bestaan uit itembanken, een reeks items (vragen) die één construct (domein) meten. Items kunnen worden geselecteerd uit een item bank om toegepast te worden als short forms (subsets van een vaststaand aantal items uit de item bank) of een zeer efficiënte computer adaptieve test (CAT). Een CAT is een computergestuurde vragenlijst waarbij, na de eerste vraag, de vervolgvragen worden geselecteerd door een computer algoritme op basis van de antwoorden op eerdere vragen.

Het Patient-Reported Outcomes Measurement Information System (PROMIS®) is het grootst op dit moment beschikbare systeem van op IRT-gebaseerde itembanken, short forms en CATs voor het meten van PROs. Voordelen van PROMIS zijn dat het ontwikkeld is gebruikmakend van een uitgebreid raamwerk en dat PROMIS instrumenten generiek zijn in plaats van ziekte-specifiek. Daarnaast is aangetoond dat PROMIS instrumenten een betere inhoudsvaliditeit en een kleinere meetfout hebben in vergelijking met traditionele ziekte-specifieke PROMs. PROMIS scores zijn bovendien makkelijk te interpreteren. Ze kunnen worden gebruikt om groepen te vergelijken in en tussen populaties. PROMIS is ontwikkeld in de Verenigde Staten (VS) en is al in meerdere talen vertaald, waaronder Nederlands-Vlaams.

Hoofdstuk 1 geeft een algemene introductie van het onderzoek in dit proefschrift en beschrijft welke stappen er nodig zijn om PROMIS te kunnen toepassen in Nederland en Vlaanderen. Het belangrijkste doel van het proefschrift is om te exploreren of PROMIS kan worden toegepast om ervaren gezondheid te meten in Nederlands-Vlaamse populaties.

De **hoofdstukken 2-6** beschrijven de validatie van de Nederlands-Vlaamse PROMIS Pijn Gedrag (v1.1), Pijn Belemmeringen (v1.1) en Lichamelijk Functioneren (v1.2) itembanken in meerdere Nederlandse en Vlaamse patiënt populaties. Alle drie de

PROMIS item banken laten goede psychometrische eigenschappen zien bij Nederlandse patiënten met chronisch pijn (**hoofdstukken 2, 3 en 4**, respectievelijk). Daarnaast laten de Nederlands-Vlaamse PROMIS Pijn Gedrag en Pijn Belemmeringen item banken goede psychometrische eigenschappen zien bij Nederlandse en Vlaamse patiënten met reumatoïde artritis (**hoofdstuk 5**). De Nederlands-Vlaamse PROMIS Lichamelijk Functioneren item bank laat ook goede resultaten zien bij Nederlandse fysiotherapie patiënten (**hoofdstuk 6**). Alle drie de item banken voldoen aan de aannames van een IRT-model en laten goede IRT-model fit zien. De item banken tonen goede cross-culturele validiteit (geen Differential Item Functioning (DIF) voor taal; Nederlands-Vlaams vs. Engels en Nederlands vs. Vlaams), afwezigheid van andere meetvariantie (geen DIF voor leeftijd, geslacht, afname modus (digitaal vs. papier) of ziekteactiviteit), goede construct validiteit, hoge betrouwbaarheid en afwezigheid van vloer- en plafondeffecten.

Naast de studies naar de psychometrische eigenschappen van de PROMIS item banken en om de validiteit van vergelijkingen tussen populaties na te gaan, zijn de item parameters van de Nederlands-Vlaamse PROMIS Pijn Gedrag, Pijn Belemmeringen en Lichamelijk Functioneren item banken vergeleken tussen verschillende klinische populaties met een aandoening van het bewegingsapparaat en een algemene populatie (**hoofdstuk 7**). De resultaten laten een paar items zien met DIF tussen de populaties, maar de impact van de DIF op de PROMIS T-scores blijkt verwaarloosbaar. Dit onderzoek levert bewijs voor de universele toepasbaarheid van de drie item banken. De item banken en hun scores kunnen worden gebruikt voor valide vergelijkingen tussen verschillende klinische populaties met een aandoening van het bewegingsapparaat, en de algemene populatie.

Hoewel het huidige PROMIS beleid is om één enkele set item parameters te gebruiken (de huidige aanbeveling is de Amerikaanse item parameters) bij het scoren van PROMIS instrumenten en in de PROMIS CAT algoritmen wereldwijd, kunnen er redenen zijn om land-specifieke item parameters te willen gebruiken in niet-Engels sprekende landen, zoals Nederland of Vlaanderen. Niet-Engelstalige PROMIS onderzoekers kunnen beargumenteren dat een populatie het meest nauwkeurig gemeten wordt met item parameters die zijn verkregen in een grote steekproef uit de betreffende populatie. Vanuit een internationaal perspectief is het echter belangrijk dat scores internationaal vergelijkbaar zijn. In **hoofdstuk 8** laat een simulatiestudie zien wat de consequenties zijn van het gebruik van verschillende sets item parameter in PROMIS CATs in niet-Engels sprekende landen. Vier opties zijn onderzocht die verschillende keuzes weerspiegelen ten aanzien van het behouden en/of herschalen (naar de Amerikaanse PROMIS meetlat) van land-specifieke item parameters. Op basis van de studieresultaten is de conclusie dat er geen evidentie is gevonden om de huidige PROMIS aanbeveling te veranderen wat betreft vergelijking op groepsniveau. CAT T-scores verkregen met behulp van Amerikaanse item parameters voor DIF-vrije items en land-specifieke item parameters, herschaald naar de Amerikaanse PROMIS meetlat, voor DIF items, of CAT T-scores verkregen met behulp van land-specifieke item parameters voor alle items, herschaald naar de Amerikaanse PROMIS meetlat, zijn vergelijkbaar met CAT T-scores verkregen

met behulp van de Amerikaanse item parameters op groepsniveau. De keuze van item parameters kan echter grotere gevolgen hebben voor individuele patiënt scores. Het gebruik van Amerikaanse of land-specifieke item parameters hebben beide hun voor- en nadelen. De PROMIS Health Organization (PHO) zal een besluit moeten nemen over voorkeurs- en toegestane regels voor het kiezen van item parameters in niet-Engels sprekende landen, om consistente rapportage van PROMIS T-scores te waarborgen over verschillende populaties en landen.

Hoofdstuk 9 geeft de algemene discussie van dit proefschrift weer, laat zien wat dit proefschrift heeft opgeleverd en geeft implicaties voor toekomstig onderzoek. Onderwerpen die belangrijk zijn voor de implementatie van PROMIS in Nederland en Vlaanderen worden besproken en een aantal voorbeelden van implementaties van PROMIS CATs in de dagelijkse klinische praktijk worden gegeven.

Dit proefschrift is een belangrijke stap in het gebruik van PROMIS in Nederland en Vlaanderen. Het is een onderdeel van een groot landelijk PROMIS initiatief. Het proefschrift draagt bij aan het bewijs dat PROMIS bestaat uit een set gestandaardiseerde PROMs van hoge kwaliteit. Het laat goede psychometrische kwaliteit zien van PROMIS item banken in meerdere patiënten-populaties. De drie onderzochte Nederlands-Vlaamse PROMIS item banken zijn geschikt en klaar voor gebruik in Nederland en Vlaanderen volgens de PROMIS-vereisten voor het vrijgeven van PROMIS instrumenten in andere talen dan het Engels en in andere landen dan Amerika. PROMIS kan goed worden toegepast in Nederland en Vlaanderen. Wereldwijd worden grote inspanningen geleverd om PROMIS te implementeren in de dagelijkse klinische praktijk en in onderzoek. Hierdoor kan ervaren gezondheid efficiënt en patiëntvriendelijk worden gemeten.