



HAL
open science

Equating and linking Patient-Reported Outcomes Measurement Information System 29-item questionnaire and 36-item Short-Form Health Survey domains using Rasch modeling

Jean-Benoît Hardouin, Joel Coste, Alain Leplège, Alexandra Rouquette

► **To cite this version:**

Jean-Benoît Hardouin, Joel Coste, Alain Leplège, Alexandra Rouquette. Equating and linking Patient-Reported Outcomes Measurement Information System 29-item questionnaire and 36-item Short-Form Health Survey domains using Rasch modeling. *Journal of Clinical Epidemiology*, 2024, 169, pp.111326. 10.1016/j.jclinepi.2024.111326 . hal-04600836

HAL Id: hal-04600836

<https://u-paris.hal.science/hal-04600836>

Submitted on 4 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ORIGINAL RESEARCH

Equating and linking Patient-Reported Outcomes Measurement Information System 29-item questionnaire and 36-item Short-Form Health Survey domains using Rasch modeling

Jean-Benoit Hardouin^{a,b}, Joel Coste^c, Alain Leplège^d, Alexandra Rouquette^{e,f,*}

^aINSERM UMR 1246, SPHERE, Methods in Patient-Centred Outcomes and Health Research, Nantes and Tours Universities, Nantes, France

^bCHU Nantes, Public Health Department, Methodology and Biostatistics Unit, Nantes, France

^cParis Cité University, Paris, France

^dDépartement d'Histoire et de Philosophie des Sciences, Laboratoire SPHERE, UMR 7219, CNRS-Université Paris Cité, Paris, France

^eUniversité Paris-Saclay, Inserm, UVSQ, CESP, Paris, France

^fAP-HP Paris-Saclay, Epidemiology and Public Health Department, Le Kremlin-Bicêtre, France

Accepted 5 March 2024; Published online 11 March 2024

Abstract

Objectives: To develop a simple, practical methodology to equate or link equivalent domains of the 36-item Short-Form Health Survey (SF-36) and the Patient-Reported Outcomes Measurement Information System 29-item questionnaire (PROMIS-29) using the Rasch framework.

Study Design and Setting: In April 2016, the PROMIS-29 and SF-36 were completed by 1501 individuals selected to be representative of the French population. For each domain common to the two questionnaires, a Partial Credit Model was fitted to the items related to that dimension in the two questionnaires. These items were then calibrated on the same metric, which enabled the scores from one questionnaire to be associated with the scores from the other.

Results: Six of the seven PROMIS-29 scales and five of the six SF-36 subscales (physical, pain, social, vitality, depression and anxiety domains) were equated or linked. Correspondence tables between scores, with a 95% confidence interval, were established for each domain. A freely available Stata macro program was developed to automatize the equating or linking process.

Conclusion: These results should facilitate comparisons across studies using the SF-36 and the PROMIS-29 in France. The equating or linking process developed is simple to implement and can be used in other countries and for other instruments. © 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

Keywords: Equating; Health-related quality of life; SF-36; PROMIS-29; Rasch modeling; Linking

1. Introduction

In recent decades, health-related quality-of-life (HRQoL) has become a key indicator of health status in epidemiology, health service research and clinical settings. Several countries, including the USA, Germany and Belgium, have implemented HRQoL surveillance using instruments such as the Center for Disease Control and Prevention “Healthy Days Measures”, the 12-Item Short-

Form Survey (SF-12) or the European Quality-of-Life 5 Dimensions [1–3]. Elsewhere, standardized instruments widely included in large population surveys have been used to estimate trends in HRQoL over time, for example, in France using the 36-item Short-Form Health Survey (SF-36) in the Medical Outcomes Study [4]. In clinical settings, systematic reviews and meta-analyses are increasingly conducted to estimate the effects of treatments or the impact of diseases on HRQoL [5]. However, the plethora of HRQoL measures now available and in use has led to serious problems of comparison and interpretation, because measures are difficult to compare, and liable to the “Tower of Babel syndrome” [6,7]. Developing common metrics using equating procedures to place HRQoL instruments on shared continuums has, therefore, become a priority to broaden the

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. INSERM U1018 - CESP, Maison de Solenn, 97 Boulevard de Port Royal, 75679, Paris Cedex 14, France.

E-mail address: alex.rouquette@gmail.com (A. Rouquette).

What is new?**Key findings**

- Subscales of the PROMIS-29 and SF-36 French version were equated or linked.
- They concerned physical, pain, social, vitality, depression, anxiety domains.

What this adds to what was known?

- We propose an equating or linking methodology that is relatively simple to implement.
- A Stata macro program to implement this methodology is freely available.

What is the implication and what should change now?

- Possible comparisons and syntheses of results from studies using these questionnaires.

common bases of measurement, and enable better monitoring and summarizing of evidence from HRQoL measures [8].

Equating HRQoL instruments or subscales has been around for at least 25 years, but there is still a long way to go to reliably document links between the main generic instruments [9]. A fundamental difference between scaling, linking, mapping and equating is that equating scores from multiple test forms assumes that these tests measure the same construct (ie, the same latent trait) though the different tests may introduce differences in difficulty [10]. To our knowledge, there has been no attempt to equate subscales of the SF-36, one of the most, if not the most widely-used generic measure assessing HRQoL and the more recent Patient-Reported Outcomes Measurement Information System 29-item questionnaire (PROMIS-29).

The objectives of the present study were to develop a simple, practical methodology to equate or to link SF-36 (V1.3 academic version) and PROMIS-29 equivalent set of items using the Rasch modeling framework, and applying it to French-language instruments. Equating was performed when all the items of the subscales from the two questionnaires to be equated were shown to be related to the same dimension, elsewhere we considered it as a linking process. Rasch models, including the partial credit model (PCM) used here, provide a framework to score several subscales simultaneously, to link them to a common scale (the latent trait), to formally test the unidimensionality of this scale and to define the score obtained with one set of items and the score obtained with another set of items as equivalent. Indeed, two properties of Rasch models are interesting to carrying out this equating or linking process: specific objectivity which allows the level of each

individual to be estimated on the latent variable using a subset of items, and the exhaustiveness property of the score on the latent trait which ensures a one-to-one relationship between each value of the score and the latent variable [11].

2. Materials and methods*2.1. Data*

We used data from 1501 subjects surveyed by telephone in April 2016 in France by an independent polling company (Ipsos) [12]. Quota sampling was used to obtain a nationally representative sample for gender, age, occupation, region, and area of residence. Socio-demographic information (age, education, income, occupation, marital status, and household size) was collected, as was the presence of selected chronic diseases diagnosed by a doctor and reported by the subject. The French versions of two HRQoL questionnaires were administered: the PROMIS-29 and the SF-36 [13,14].

2.1.1. PROMIS-29

The PROMIS-29 is a generic patient-reported outcome measure developed within the PROMIS Initiative, a cooperative group of research centers implementing mixed-method development processes to create domain-specific, person-centered measures of physical, mental and social health for use across diseases [15–18]. It includes 28 items, four from each of the seven PROMIS domains (depression, anxiety, physical function, pain interference, fatigue, sleep disturbance, and ability to participate in social roles and activities), and a single item on pain intensity. Each item has five response options (scored 1 to 5), except for pain intensity, which has eleven response options (scored 0 to 10). A total raw score is computed for each domain as the sum of the responses to each question within the domain, yielding seven domain scores, each scored from 4 to 20.

2.1.2. SF-36

The SF-36 is the most widely-used generic health status measure, with 36 items exploring eight domains (physical function, role physical, pain, general health, vitality, social function, role emotional, mental health) with scores ranging from 0 (worst perceived status) to 100 on each dimension, as described in the MOS-SF36 user's guide [13,19,20].

2.2. Statistical analyses

Stata software version 18 was used for all analyses except for Rasch model fit, which was assessed using RUMM2030 software and Confirmatory Factor Analyses (CFA) performed using MPlus version 8.10 [21–23].

2.2.1. Dimensionality

First, the four authors examined the items of the subscales of the two questionnaires to identified candidates for equating given the similarity of the concept they are intended to assess (Table 1). Then, a unidimensional CFA model was fitted on all the items from the PROMIS-29 and SF-36 dimensions to be equated to check for unidimensionality. The response coding for items from the PROMIS-29 dimensions pain interference, fatigue, anxiety and depression was reversed so that a higher score indicated better status, as in the SF-36. As local independence is required to apply Rasch models, three new items were created to replace each of the three sets of items that were not locally independent in the physical function dimension of the SF-36 (Supplementary material 1). The robust weighted least squares estimator, appropriate for categorical data, was used for CFA and residual covariance between SF-36 items were allowed [23]. Hu et al. cut-off criteria based on the Root Mean Square Error Approximation (RMSEA), the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) were examined to evaluate the fit of CFA models [24]. In case of poor CFA fit, an exploratory factorial analysis (EFA) was applied to all the items in the PROMIS-29 and SF-36 dimensions to be equated in order to explore the possibilities to link these dimensions.

2.2.2. Equating or linking

For each dimension, once a unidimensional set of items was found, two raw scores were computed using items from the PROMIS-29 and from the SF-36 separately. Estimates of the latent trait and its 95% confidence interval (95% CI) were produced for each value of these raw scores using a PCM. Then, for each of the three estimates (latent trait, 95% CI lower and upper bounds) associated with each value of the raw score on one questionnaire, we searched for the corresponding value in terms of raw score on the other questionnaire using linear interpolation rounded to the closest integer. Finally, all raw score values for the two questionnaires were linearly transformed to identify their equivalent on the classic score ranges of the questionnaires (4–20 for PROMIS-29 and 0–100 for SF-36). Previously-reversed PROMIS dimensions were reversed once more to be in line with the classic PROMIS-29 scores used in practice.

Pearson's correlation coefficients between the classic SF-36 score and the SF-36 raw score used to conduct the equating (or linking) were estimated as the SF-36 classic scoring was not used for some of the dimensions (new items, scoring weights not considered, etc.). They were also assessed between the classic score of each questionnaire considering that a Pearson's coefficient < 0.6 would identify dimensions that were insufficiently correlated to be equated. Lastly, Intraclass Correlation Coefficient (ICC) was assessed between the classic score on one questionnaire and the equated value (central point of the 95% CI), plausible equated values (randomly drawn within the 95%

CI) and the equated value obtained by the widely used equipercentile method on the same questionnaire corresponding to the classic score of the other questionnaire [25]. Agreement between classic score and equated value was considered as poor, fair, good or excellent if the ICC was below 0.4, between 0.4 and 0.59, between 0.6 and 0.74 and above 0.74 respectively [26]. A comparison between the Rasch equating process developed in this study and the equipercentile method was also conducted. This involved estimating the proportion of individuals with an absolute difference between the equated and the classic scores exceeding 10 and 20 points for the SF36 and 2 and 4 points for the PROMIS-29.

3. Results

The characteristics of the 1501 participants in the survey are shown in Table 2. The mean age was 48.3 years (SD = 17.0), 52% were female, 49% had higher education, 68% were married or living with a partner, 52% had an occupation and 38% reported a chronic disease.

3.1. Domains to be equated

As shown in Table 1, only five SF-36 dimensions were identified as candidates to be equated to six PROMIS-29 dimensions: the physical (Physical Function for SF36 and PROMIS-29), pain (Bodily Pain SF-36 – Pain Interference PROMIS-29), social (Social Function SF-36 – Social Participation PROMIS-29), vitality (Vitality SF-36 – Fatigue PROMIS-29), depression (Mental Health SF-36 – Depression PROMIS-29), and anxiety (Mental Health SF-36 – Anxiety PROMIS-29) dimensions. Depression and anxiety items were grouped in the SF-36 Mental Health dimension while they were found in two dimensions in the PROMIS-29.

The unidimensionality assumption was supported for the physical (using three items and three new items from the SF-36, and four PROMIS-29 items, RMSEA = 0.086 [0.077; 0.096], CFI = 0.993, TLI = 0.985), the pain (using two SF-36 items and four PROMIS-29 items, RMSEA = 0.063 [0.047; 0.079], CFI = 1.000, TLI = 0.999) and the social (using two SF-36 items and four PROMIS-29 items, RMSEA = 0.053 [0.038; 0.069], CFI = 0.999, TLI = 0.998) dimensions. Equating was, thus considered possible for these three domains.

For the vitality dimension, as the unidimensionality assumption was not met, an EFA was applied to the four SF-36 vitality items and the four PROMIS-29 fatigue items, yielding a 2-factor solution with the two positively worded SF-36 items (with response coding reversed) loading on the second factor. It was, thus decided to link the raw score computed on the two negatively worded SF-36 items (VT3 and VT4) to the raw score of the 4-item PROMIS-

Table 1. Identification of the candidate dimensions of the PROMIS-29 and the SF-36 to be equated

| SF36 | | PROMIS-29 | |
|--|---|---|---|
| Physical function (PF) 10 items 3 response options | SF-PF1 - Vigorous activities, such as running... ^a SF-PF2 - Moderate activities, such as moving a table... ^a SF-PF3 - Lifting or carrying groceries SF-PF4 - Climbing several flights of stairs ^a SF-PF5 - Climbing one flight of stairs ^a SF-PF6 - Bending, kneeling or stooping SF-PF7 - Walking more than a mile ^a SF-PF8 - Walking several blocks ^a SF-PF9 - Walking one block ^a SF-PF10 - Bathing or dressing yourself | Physical function (PF) 4 items 5 response options | PR-PF1 - Are you able to do chores such as vacuuming... PR-PF2 - Are you able to go up and down stairs... PR-PF3 - Are you able to go for a walk of at least ... PR-PF4 - Are you able to run errands and shop? |
| Role physical 4 items 2 response options | SF-RP1 - Cut down the amount of time you spent on work... SF-RP2 - Accomplished less than you would like SF-RP3 - Were limited in the kind of work or other activities SF-RP4 - Had difficulty performing the work or other ... | | |
| Bodily pain 2 items 6 (BP1) and 5 (BP2) response options | SF-BP1 - How much bodily pain have you had during ... SF-BP2 - ..., how much did pain interfere with your normal work... | Pain interference ^c 4 items 5 response options | PR-PI1 - ... interfere with your day-to-day activities? PR-PI2 - ... interfere with work around the home? PR-PI3 - ... interfere with your ability to participate in social... PR-PI4 - ... interfere with your household chores? |
| | | Pain intensity 1 item 11 response options | PR-PI5 - How would you rate your pain on average? |
| General health 4 items 5 response options | SF-GH1 - I seem to get sick a little easier than other people SF-GH2 - I am as healthy as anybody I know SF-GH3 - I expect my health to get worse SF-GH4 - My health is excellent | | |
| Vitality 4 items 6 response options | SF-VT1 - Did you feel full of pep? ^b SF-VT2 - Did you have a lot of energy? ^b SF-VT3 - Did you feel worn out? SF-VT4 - Did you feel tired? | Fatigue ^c 4 items 5 response options | PR-FA1 - I feel fatigued PR-FA2 - I have trouble starting things because I am tired PR-FA3 - How run-down did you feel on average? PR-FA4 - How fatigued were you on average? |
| Social function 2 items 6 response options | SF-SF1 - ... to what extent has your physical health or emotional problems interfered with your normal social... SF-SF2 - ... how much of the time has your physical health or emotional problems interfered with your social activities... | Social participation 4 items 5 response options | PR-SP1 - ... of my regular leisure activities with others PR-SP2 - ... of the family activities that I want to do PR-SP3 - ... of my usual work (include work at home) PR-SP4 - ... of the activities with friends that I want to do |
| Mental Health 5 items 6 response options | SF-MH1 - Have you been a very nervous person? SF-MH2 - Have you felt so down in the dumps that... SF-MH3 - Have you felt calm and peaceful? ^b SF-MH4 - Have you felt downhearted and blue? SF-MH5 - Have you been a happy person? ^b | Depression ^c 4 items 5 response options Anxiety ^c 4 items 5 response options | PR-DE1 - I felt worthless PR-DE2 - I felt helpless PR-DE3 - I felt depressed PR-DE4 - I felt hopeless PR-AN1 - I felt fearful PR-AN2 - I found it hard to focus on anything other... PR-AN3 - My worries overwhelmed me PR-AN4 - I felt uneasy |

(Continued)

Table 1. Continued

| SF36 | | PROMIS-29 |
|--|---|---|
| Role emotional 3 items 2 response options | SF-RE1 - Cut down the amount of time you spent on work... SF-RE2 - Accomplished less than you would like SF-RE3 - Didn't do work or other activities as carefully... | |
| Health transition 1 item 5 response options | SF-HT - Compared to 1 yr ago, how would you rate your health in general now? | |
| | Sleep disturbance 4 items 5 response options | PR-SD1 - My sleep quality was PR-SD2 - My sleep was refreshing PR-SD3 - I had a problem with my sleep PR-SD4 - I had difficulty falling asleep |

^a Sets of items not locally independent for which new items equal to the sum of the former item were created ([supplementary material 1](#)).

^b Positively worded items that did not load on the same factor as the negatively worded items in the exploratory factor analysis, even when the response coding was reversed. These items have not been used in the analysis.

^c A higher score indicates poorer status on the dimension, in gray: no dimension to be equated in the questionnaire.

29 fatigue dimension (RMSEA = 0.082 [0.067; 0.098], CFI = 0.999, TLI = 0.998).

For the mental health dimension, the unidimensionality assumption was not met, whether the PROMIS-29 dimension considered was depression or anxiety. An EFA was, thus applied to the five SF-36 items in the mental health dimension first with the four PROMIS-29 items in the depression dimension, followed by the four PROMIS-29 items in the anxiety dimension. In both cases, it showed a 2-factor solution with the two positively-worded SF-36 items (response coding reversed) loading on the second factor. It was, thus decided to link the raw score computed on the three negatively-worded SF-36 items (MH1, MH2 and MH4) with the raw score of the 4-item PROMIS-29 depression dimension (RMSEA = 0.086 [0.073; 0.099], CFI = 0.997, TLI = 0.994) and with the raw score of the 4-item PROMIS-29 anxiety dimension separately (RMSEA = 0.062 [0.049; 0.076], CFI = 0.998, TLI = 0.996).

3.2. Equated or linked scores

Figure 1 is a graphic presentation of the equating process, taking the physical dimension as an example. The estimation of the latent trait level corresponding to an SF-36 classic score of 70 on the physical functioning dimension was -2.67 [-3.67; -1.66] (red line). This latent trait estimation corresponded to 18 [15; 19] on the PROMIS-29 physical functioning classic scale. Conversely, the estimation of the latent trait level corresponding to a PROMIS-29 physical functioning classic score of 10 was -4.75 [-5.65; -3.85] (green line). This latent trait estimation corresponded to 23 [10; 42] on the SF-36 physical functioning classic scale.

Figure 2 (equated domains: physical, pain and social dimensions) and 3 (linked domains: vitality, depression and anxiety dimensions) show the relationships between SF-

36 classic scores and PROMIS-29 classic scores with their 95% CIs, for each dimension ([Supplementary Tables 1 and 2](#)).

As shown in [Table 3](#), Pearson's correlations between the raw scores used for equating and the classic SF-36 scores were almost perfect (>0.99) for the equated domains and very high (>0.89) for the linked domains. Pearson's coefficients between the classic scores on the two questionnaires ranged from 0.62 (social) to 0.75 (pain and fatigue), which seemed compatible with the scope for equating or linking. A good to excellent agreement was found between the classic SF-36 score and the SF-36 equated value (ICC ranged from 0.63 to 0.77) and good to excellent agreement between the classic PROMIS-29 score and the PROMIS-29 equated value (ICC ranged from 0.61 to 0.78). These values were slightly lower when computed with the equipercetile equated values or with plausible equated values randomly drawn within the 95% CI.

On average, across the six domains, for 60% (respectively 81%) of the sample, the absolute difference between the equated and the classical score was 10 points or lower (respectively 20 points) for the SF36 with the proposed method (50% and 80% with the equipercetile method). These proportions were 74% (respectively 91%) for an absolute difference of 2 points or lower (respectively 4 points) concerning PROMIS-29 (74% and 91% with the equipercetile method).

4. Discussion

This study reports the first and successful attempt to equate the PROMIS-29 and the SF-36 (V1.3 academic version) measures for most of their subscales. Equating was possible for physical, pain, and social domains but not for fatigue, anxiety and depression (PROMIS-29

Table 2. Socio-demographic characteristics of subjects ($N = 1501$)

| Characteristics | N (%) |
|---|----------|
| Gender | |
| Men | 725 (48) |
| Women | 776 (52) |
| Age | |
| 18–24 | 279 (19) |
| 25–34 | 244 (16) |
| 35–44 | 277 (18) |
| 45–54 | 261 (17) |
| 55–64 | 214 (15) |
| 65–74 | 226 (15) |
| ≥75 | 279 (19) |
| Region | |
| Ile-de-France | 280 (19) |
| North-West | 338 (23) |
| North-East | 337 (23) |
| South-West | 170 (11) |
| South-East | 376 (25) |
| Residence area | |
| Rural | 345 (23) |
| Fewer than 20,000 inhabitants | 258 (17) |
| 20,000–99,999 inhabitants | 204 (14) |
| 100,000 inhabitants and more | 448 (30) |
| Paris area | 246 (16) |
| Current occupation | |
| Managers and Professionals | 157 (11) |
| Technicians, Clerks, Service workers | 332 (23) |
| Workers, self-employed, Armed forces | 259 (18) |
| Inactive/Unemployed | 703 (48) |
| Missing | 50 |
| Education | |
| Primary school or less | 114 (8) |
| Secondary (professional) | 357 (24) |
| Secondary (general) | 291 (19) |
| University – Higher education (short cycle) | 314 (21) |
| University - (medium cycle) | 219 (14) |
| University - (long cycle) | 206 (14) |
| Marital Status | |
| Never Married (Single) | 309 (21) |
| Living with a partner | 223 (15) |
| Married/Civil Partnership | 796 (53) |
| Separated | 29 (2) |
| Divorced | 104 (7) |
| Widowed | 40 (3) |

dimensions) with vitality and mental health (SF-36 dimensions). In these latter, a linking process was nevertheless proposed while it was impossible for sleep disturbance, which has no counterpart in the SF-36, and for general health, which has no counterpart in the PROMIS-29. We provide [Figures 2 and 3](#) and supplementary tables that

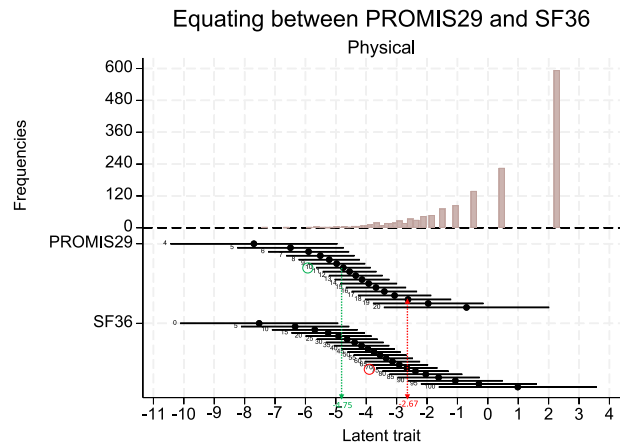


Figure 1. The equating process, with the physical domain as an example. The X-axis represents the latent trait, and the estimated latent trait values (circles) and 95% confidence intervals (lines) for each score on the SF-36 (ranging 0–100) are shown at the bottom, while these values for each classic score on the PROMIS-29 (ranging from 4 to 20) are shown in the middle of the graph. At the top are represented the frequencies according to the latent trait in the sample. Red line: the estimation of the latent trait level corresponding to an SF-36 classic score of 58 on the physical functioning dimension was -2.67 [-3.67 ; -1.66] which corresponded to 18 [15; 19] on the PROMIS-29 physical functioning classic scale. Green line: the estimation of the latent trait level corresponding to a PROMIS-29 physical functioning classic score of 10 was -4.75 [-5.65 ; -3.85] which corresponded to 23 [10; 42] on the SF-36 physical functioning classic scale. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article).

can be easily used in practice when one of these two questionnaires is used and when a comparison is needed with previously-collected data using the other questionnaire. For example, if the PROMIS-29 were to be chosen to evaluate HRQoL among individuals enrolled in a cohort study for its good psychometric properties, [Supplementary Table 2](#) could be used to equate each value of the PROMIS-29 score to SF-36 metrics for each individual if a historic comparison is required with data previously collected using the SF-36. The 95% CIs were provided to enable users wishing to take the precision of the estimation into account and avoid overestimation of equated SF-36 score variance. In the Appendix, we provide a Stata macro program that automatizes the equating process presented in this paper, freely available (command “`ssc install pcm`” directly from Stata).

Equating the pain dimensions (4 PROMIS-29 items and 2 SF-36 items) was simple, with the shape of the curve of the equated scores close to linear and good to excellent agreement between classic score and equated value (ICC from 0.68 to 0.78). This finding is not really surprising in view of a previously published study by Cook et al. equating SF-36 bodily pain with the PROMIS pain interference bank (41 items) [27]. Equating physical dimensions (4 PROMIS-29 items and 10 SF-36 items) was also relatively simple, only requiring the creation of new

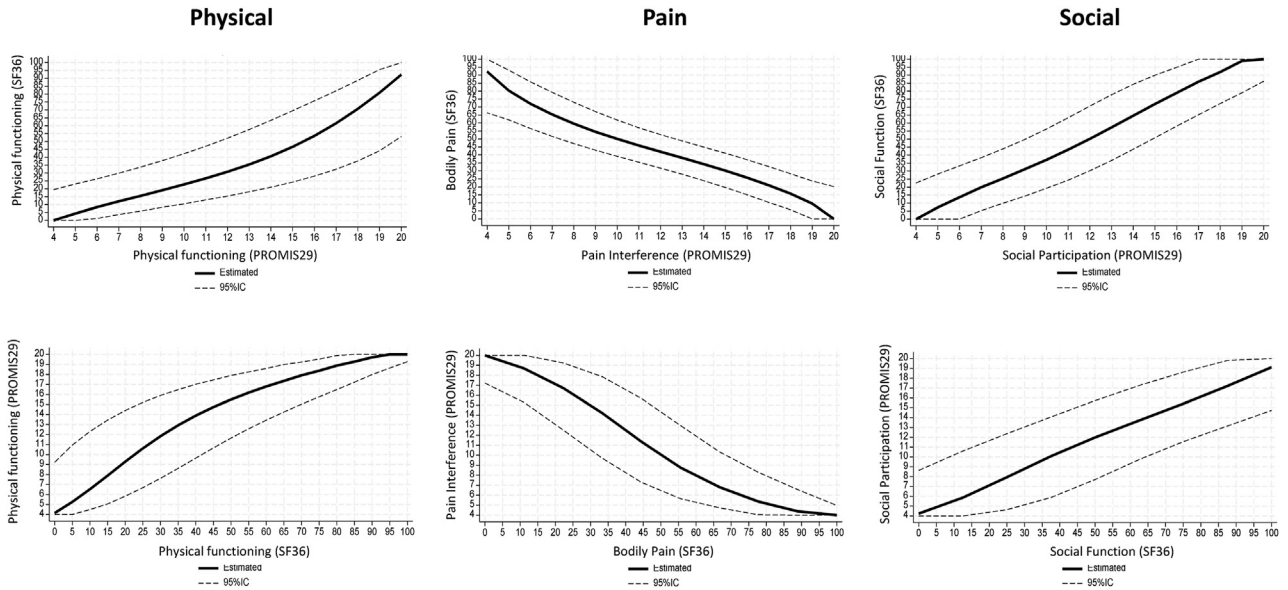


Figure 2. Equated SF-36 scores with PROMIS-29 scores (top, continuous line) and equated PROMIS-29 scores with SF-36 scores (bottom, continuous line) with their 95% confidence intervals (dotted lines) for the physical, pain and social domains.

items for locally dependent SF-36 items. This result is in line with previous successful attempts to equate SF-36 physical functioning and HAQ scales to the PROMIS physical functioning bank (76 items) [28]. However, the shape of the curve of the equated scores also evidences the already known ceiling effect that characterizes PROMIS-29 dimensions [12,29]. Similar satisfactory equating was obtained for the social dimension (4 PROMIS-29 items and 2 SF-36 items).

Linking the PROMIS-29 and SF-36 vitality-fatigue and mental health dimensions was more difficult, and required dropping two SF-36 items from these domains to obtain acceptable model fits. This was hardly surprising, as well-known dimensionality problems affect the SF-36 vitality and mental health scales, and also because of the apparent

absence of reported attempts to link these scales to the PROMIS banks, contrary to what happened with other well-known measurement instruments for anxiety or depression [30–33]. Despite these difficulties, linking fatigue and vitality, anxiety and mental health and depression and mental health was found satisfactory for the curves of the linked scores (close to linear in all cases) and for ICC between classic scores and expected values (good to excellent agreement). However, this linking process is less straightforward than for other domains because the classic SF-36 score cannot be used directly. Due to dimensionality issues affecting the SF36 mental health and vitality scales, a derived score must be calculated before linking can be applied. Furthermore, this process may yield different derived SF-36 mental health scores depending on whether

Table 3. Pearson’s correlation coefficients (PCCs) and Intraclass correlation coefficients (ICCs) between the SF-36 and PROMIS-29 classic scores and scores used in the equating or linking process

| Domains | SF36 classic score | | | | PROMIS29 classic score | | | |
|-----------------|-----------------------|-------------------------------|--|--|---|--|---|---|
| | SF-36 raw score (PCC) | PROMIS-29 classic score (PCC) | SF-36 Rasch-equated or linked value ^a (ICC) | SF-36 equipercentile-equated or linked value (ICC) | SF-36 Rasch-equated or linked plausible values ^b (ICC) | PROMIS-29 Rasch-equated or linked value ^a (ICC) | PROMIS-29 equipercentile-equated or linked values (ICC) | PROMIS-29 Rasch-equated or linked plausible values ^b (ICC) |
| Equated domains | | | | | | | | |
| Physical | 1.00 | 0.63 | 0.66 | 0.66 | 0.37 | 0.63 | 0.62 | 0.59 |
| Pain | >0.99 | −0.75 | 0.77 | 0.65 | 0.68 | 0.78 | 0.75 | 0.69 |
| Social | 1.00 | 0.62 | 0.63 | 0.62 | 0.58 | 0.61 | 0.61 | 0.48 |
| Linked domains | | | | | | | | |
| Fatigue | 0.89 ^c | −0.75 | 0.76 | 0.72 | 0.71 | 0.75 | 0.73 | 0.67 |
| Depression | 0.93 ^d | −0.74 | 0.72 | 0.65 | 0.62 | 0.72 | 0.70 | 0.64 |
| Anxiety | 0.93 ^d | −0.70 | 0.69 | 0.66 | 0.61 | 0.69 | 0.67 | 0.59 |

^a central point of the 95% CI.

^b randomly drawn within the 95% CI.

^c raw score computed using two of the four items from the SF-36 Vitality dimension.

^d raw score computed using three of the five items from the SF-36 Mental Health dimension.

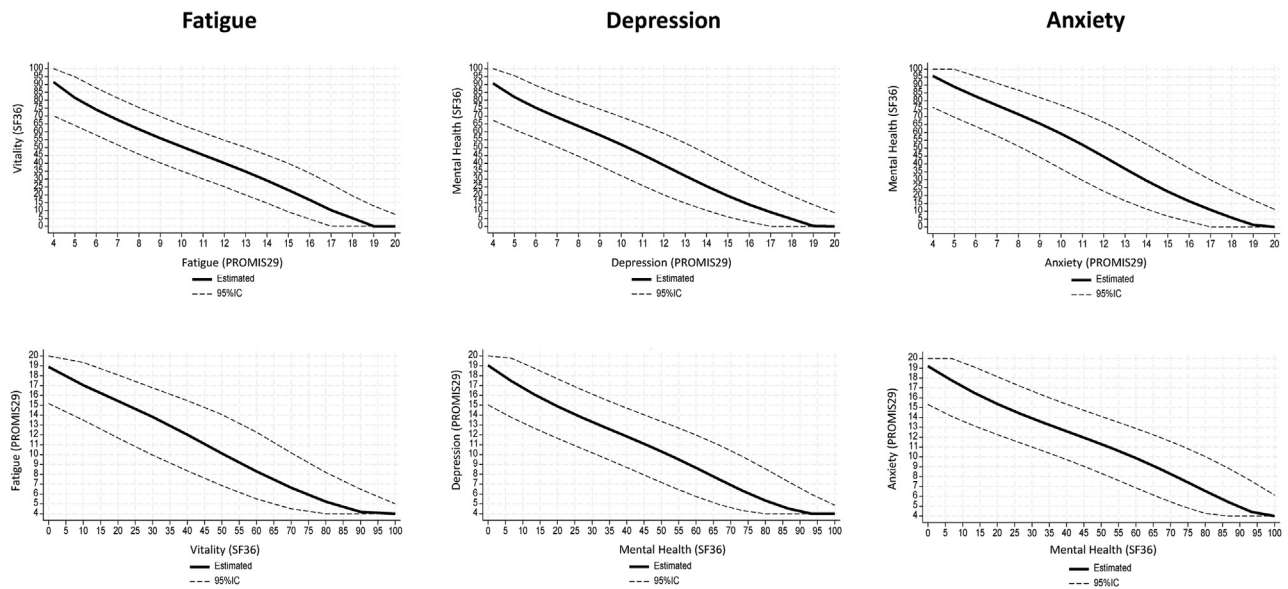


Figure 3. Linked SF-36 scores with PROMIS-29 scores (top, continuous line) and linked PROMIS-29 scores with SF36 scores (bottom, continuous line) with their 95% confidence intervals (dotted lines) for the fatigue, depression and anxiety domains.

one is looking at the anxiety or depression dimension, which are not distinct in the SF-36 mental health scale.

The method used in this study can easily be replicated in other countries, provided that a sample of participants who completed the two questionnaires to be equated is available. Nevertheless, it could yield different results in other countries due to transcultural disparities. Additionally, a similar, if not better, agreement (ICC) was found between the equated and classic scores than for the equipercentile method. However, further studies should explore the performance of these methods in case of missing data as we can anticipate better performance for the Rasch based method.

One limit of our study is the use of general population data that covers only about half of the score range to be equated (Fig 1) leading to lower precision for equated values in the other half. However, compared to classical linking methods (as equipercentile linking), latent variable-based approaches (Rasch models, IRT), as we used in this study, provide results that are not dependent of the sample, therefore, results can be easily generalized [34]. Another limit is that our method requires a unidimensional set of local independent items and this led to adaptations of original scales that reduce the ease with which our method can be applied.

5. Conclusions

Overall, six of the seven PROMIS-29 scales and five of the six SF-36 subscales were equated or linked. This should enable comparisons across studies, and also the synthesis of results for numerous studies, surveys or clinical trials using these instruments. The methodology presented in this paper is relatively simple to implement (especially with the freely

available Stata macro program provided in the Appendix) and can, therefore, be used in other countries and for other HRQoL scales.

CRedit authorship contribution statement

Jean-Benoit Hardouin: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Joel Coste:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Alain Leplège:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Alexandra Rouquette:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

None.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2024.111326>.

References

- [1] Zahran HS, Kobau R, Moriarty DG, Zack MM, Holt J, Donehoo R, et al. Health-related quality of life surveillance—United States, 1993–2002. *MMWR Surveill Summ* 2005;54:1–35.
- [2] Andersen HH, MuÈhlbacher A, NuÈbling M, Schupp J, Wagner GG. Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Schmollers Jahrb* 2007;127:172–82.
- [3] Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J. Methodological basics and evolution of the Belgian health interview survey 1997–2008. *Arch Public Health* 2013;71:24.
- [4] Clause-Verdreau A-C, Audureau È, Leplège A, Coste J. Contrasted trends in health-related quality of life across gender, age categories and work status in France, 1995–2016: repeated population-based cross-sectional surveys using the SF-36. *J Epidemiol Community Health* 2019;73:65–72.
- [5] Johnston BC, Thorlund K, Schünemann HJ, Xie F, Murad MH, Montori VM, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health Qual Life Outcomes* 2010;8:116.
- [6] Bjorner JB. Solving the tower of Babel problem for patient-reported outcome measures. *Psychometrika* 2021;86:747–53.
- [7] Kaplan RM, Hays RD. Health-related quality of life measurement in public health. *Annu Rev Public Health* 2022;43:355–73.
- [8] Fisher WP. Physical disability construct convergence across instruments: towards a universal metric. *J Outcome Meas* 1997;1:87–113.
- [9] Fisher WP, Eubanks RL, Marier RL. Equating the MOS SF36 and the LSU HSI physical functioning scales. *J Outcome Meas* 1997;1:329–62.
- [10] Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking: Methods and Practices*. New York, NY: Springer Science & Business Media; 2014.
- [11] Bond T. *Applying the rasch model: fundamental measurement in the human sciences*, third edition. 3rd ed. New York: Routledge; 2015.
- [12] 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. *Qual Life Res* 2018;27:999–1014.
- [13] Leplège A, Ecosse E, Coste J, Pouchot J, Perneger T. Le questionnaire MOS SF-36: manuel de l'utilisateur et guide d'interprétation des scores. *Estem* 2001;12:137–49.
- [14] Coste J, Rouquette A, Valderas JM, Rose M, Leplège A. The French PROMIS-29. Psychometric validation and population reference values. *Rev Epidemiol Sante Publique* 2018;66:316–23.
- [15] DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care* 2007;45:S12–21.
- [16] Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, et al. The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. *Health Qual Life Outcomes* 2013;11:210.
- [17] 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). *Med Care* 2007;45:S22–31.
- [18] 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. *J Clin Epidemiol* 2010;63:1179–94.
- [19] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- [20] Leplège A, Ecosse E, Verdier A, Perneger TV. The French SF-36 health survey: translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol* 1998;51:1013–23.
- [21] StataCorp LP. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp, L.P.; 2015.
- [22] Andrich D, Sheridan BS, Luo G. *Rumm2030: Rasch Unidimensional Measurement Models* [computer software]. Perth, Western Australia: RUMM Laboratory; 2010.
- [23] Muthén LK, Muthén BO. (1998–2012). *MPlus. Statistical analysis with latent variables. User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén; 2012.
- [24] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.
- [25] Muraki E, Hombo CM, Lee Y-W. Equating and linking of performance assessments. *Appl Psychol Meas* 2000;24:325–37.
- [26] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284–90.
- [27] Cook KF, Schalet BD, Kallen MA, Rutsohn JP, Cella D. Establishing a common metric for self-reported pain: linking BPI pain interference and SF-36 bodily pain subscale scores to the PROMIS pain interference metric. *Qual Life Res* 2015;24:2305–18.
- [28] Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a common metric for physical function: linking the HAQ-DI and SF-36 PF subscale to PROMIS® physical function. *J Gen Intern Med* 2015;30:1517–23.
- [29] Garratt AM, Coste J, Rouquette A, Valderas JM. The Norwegian PROMIS-29: psychometric validation in the general population for Norway. *J Patient Rep Outcomes* 2021;5:86.
- [30] 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. *J Anxiety Disord* 2014;28:88–96.
- [31] Coste J, Bouée S, Ecosse E, Leplège A, Pouchot J. Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. *Qual Life Res* 2005;14:641–54.
- [32] 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. *Psychol Assess* 2014;26:513–27.
- [33] Kaat AJ, Newcomb ME, Ryan DT, Mustanski B. Expanding a common metric for depression reporting: linking two scales to PROMIS® depression. *Qual Life Res* 2017;26:1119–28.
- [34] Council of Chief State School Officers. *A practitioner's introduction to equating with primers on classical test theory and item response theory*, Revised First Edition. Washington, CA: Council of Chief State School Officers; 2018.