

RESEARCH

Open Access



# Converting PROMIS<sup>®</sup>-29 v2.0 profile data to SF-36 physical and mental component summary scores in patients with cardiovascular disorders

Gregor Liegl<sup>1\*</sup>, Felix H. Fischer<sup>1</sup>, Carl N. Martin<sup>2,26,27</sup>, Maria Rönnefarth<sup>2</sup>, Annelie Blumrich<sup>2</sup>, Michael Ahmadi<sup>2,3,4</sup>, Leif-Hendrik Boldt<sup>6,7,14</sup>, Kai-Uwe Eckardt<sup>3,5</sup>, Matthias Endres<sup>3,4,7,8,9</sup>, Frank Edelmann<sup>2,6,7</sup>, Holger Gerhardt<sup>2,7,10</sup>, Ulrike Grittner<sup>2,11</sup>, Arash Haghikia<sup>2,7,14</sup>, Norbert Hübner<sup>2,7,12,13</sup>, Ulf Landmesser<sup>2,7,14</sup>, David Leistner<sup>15,28,29</sup>, Knut Mai<sup>7,16,17,23</sup>, Jil Kollmus-Heege<sup>2,11</sup>, Dominik N. Müller<sup>2,7,10,13,18</sup>, Christian H. Nolte<sup>2,3,4,7</sup>, Sophie K. Piper<sup>2,11,24</sup>, Kai M. Schmidt-Ott<sup>2,5,12,25</sup>, Tobias Pischon<sup>2,7,13,21,22</sup>, Simrit Rattan<sup>2,11</sup>, Ira Rohrpasser-Napierkowski<sup>2</sup>, Katharina Schönath<sup>2</sup>, Jeanette Schulz-Menger<sup>7,12</sup>, Oliver Schweizerhof<sup>2,11</sup>, Joachim Spranger<sup>2,7,15,16,23</sup>, Joachim E. Weber<sup>2,3,4,7</sup>, Martin Witzernath<sup>19,20</sup>, Sein Schmidt<sup>2,3,4†</sup> and Matthias Rose<sup>1†</sup>

## Abstract

**Background** Health-related quality of life (HRQL) has become an important outcome parameter in cardiology. The MOS 36-Item Short-Form Health Survey (SF-36) and the PROMIS-29 are two widely used generic measures providing composite HRQL scores. The domains of the SF-36, a well-established instrument utilized for several decades, can be aggregated to physical (PCS) and mental (MCS) component summary scores. Alternative scoring algorithms for correlated component scores (PCS<sub>c</sub> and MCS<sub>c</sub>) have also been suggested. The PROMIS-29 is a newer but increasingly used HRQL measure. Analogous to the SF-36, physical and mental health summary scores can be derived from PROMIS-29 domain scores, based on a correlated factor solution. So far, scores from the PROMIS-29 are not directly comparable to SF-36 results, complicating the aggregation of research findings. Thus, our aim was to provide algorithms to convert PROMIS-29 data to well-established SF-36 component summary scores.

**Methods** Data from  $n = 662$  participants of the Berlin Long-term Observation of Vascular Events (BeLOVE) study were used to estimate linear regression models with either PROMIS-29 domain scores or aggregated PROMIS-29 physical/mental health summary scores as predictors and SF-36 physical/mental component summary scores as outcomes. Data from a subsequent assessment point ( $n = 259$ ) were used to evaluate the agreement between empirical and predicted SF-36 scores.

<sup>†</sup>Sein Schmidt, Matthias Rose these authors contributed equally.

\*Correspondence:  
Gregor Liegl  
gregor.liegl@charite.de

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Results** PROMIS-29 domain scores as well as PROMIS-29 health summary scores showed high predictive value for PCS, PCS<sub>c</sub> and MCS<sub>c</sub> ( $R^2 \geq 70\%$ ), and moderate predictive value for MCS ( $R^2 = 57\%$  and  $R^2 = 40\%$ , respectively). After applying the regression coefficients to new data, empirical and predicted SF-36 component summary scores were highly correlated ( $r > 0.8$ ) for most models. Mean differences between empirical and predicted scores were negligible ( $|SMD| < 0.1$ ).

**Conclusions** This study provides easy-to-apply algorithms to convert PROMIS-29 data to well-established SF-36 physical and mental component summary scores in a cardiovascular population. Applied to new data, the agreement between empirical and predicted SF-36 scores was high. However, for SF-36 mental component summary scores, considerably better predictions were found under the correlated (MCS<sub>c</sub>) than under the original factor model (MCS). Additionally, as a pertinent byproduct, our study confirmed construct validity of the relatively new PROMIS-29 health summary scores in cardiology patients.

**Keywords** PROMIS-29, SF-36, Health composite scores, Cardiovascular diseases, Mapping, Patient-reported outcomes, Outcome measures, Health-related quality of life

## Background

The assessment of health-related quality of life (HRQL) is becoming increasingly important when it comes to evaluating and improving healthcare in many medical fields, including cardiology [1–3]. Regulatory bodies worldwide, such as the European Medicines Agency and the U.S. Food and Drug Administration, have recommended measuring HRQL for several years to evaluate the efficacy and safety of medical treatments [4]. As a consequence, many different HRQL measurement instruments have been developed and used in patients with cardiovascular disease [5]. Since the results of different HRQL measures cannot be directly compared, the aggregation of research findings is often difficult or not possible at all [6]. Therefore, there is an urgent need for developing methods that allow to convert the scores of one HRQL measure into the scores of another [7, 8].

HRQL is a multidimensional construct most commonly assessed using patient-reported outcome measures (PROM) [9]. Over the past decades, a vast number of PROMs have been developed for the assessment of many different domains of physical and psychosocial health, such as physical functioning, pain intensity and interference, fatigue, sleep disturbance, depression, anxiety, and many more [10]. The use of such narrowly specified health domains has the advantage that outcome assessments can be adapted to specific contexts in the best possible way [11]. However, for certain research questions, it appears to be more meaningful to use composite measures that combine different aspects of HRQL into an aggregated score representing a broader health concept [12]. This may be particularly the case when a general indicator of physical or mental health is required to compare groups from different populations, or when the population of interest is heterogeneous and has a wide range of impaired HRQL domains [13].

The 36-Item Short-Form Health Survey (SF-36), developed in the Medical Outcome Study in the early 1990s, is

still one of the most frequently used generic HRQL measures [12, 14, 15]. The SF-36 consists of eight domains which can be scored separately. In addition, the individual subscale scores of each domain can be aggregated to physical (PCS) and mental (MCS) component summary scores (i.e., weighted sum scores), which are widely used composite measures of physical and mental health [15, 16]. These two distinct higher-order summary scores were derived from principal component analysis, explaining more than 80% of reliable variance of the eight SF-36 subscales [16].

Originally, PCS and MCS were derived using an orthogonal factor model, meaning that PCS and MCS were assumed to be uncorrelated when establishing scoring algorithms. This approach has some advantages over an oblique (i.e. correlated) factor model, including simplicity and straightforward interpretation of the individual scales [16]. Nonetheless, subsequent research has shown that the assumption that physical and mental health are independent constructs may not hold [17–19]. As a consequence, modified scoring algorithms for correlated SF-36 summary scores (PCS<sub>c</sub> and MCS<sub>c</sub>) have been suggested [20]. Although many studies have shown that mental and physical health are actually quite strongly related, making correlated components more plausible than uncorrelated components [20, 21], PCS<sub>c</sub> and MCS<sub>c</sub> are still used less frequently than original SF-36 PCS and MCS.

The 29-item Patient-Reported Outcomes Measurement Information System (PROMIS) adult profile measure (PROMIS-29) is a newer generic measure of HRQL which is increasingly used as an alternative to the SF-36 [10, 22]. The PROMIS-29 assesses eight domains related to physical and psychosocial health, which slightly differ from the domains of the SF-36. However, the largest difference – and advantage – of the PROMIS-29 is that its items were included for original calibration of comprehensive PROMIS item banks using item response

theory (IRT) methodology [23]. Thus, PROMIS-29 domain scores are placed on the same T-score metric as all other items of the corresponding domain-specific item bank. Analogous to the SF-36, physical and mental health summary scores can be derived from PROMIS-29 domain scores [22]; respective scoring algorithms are based on a correlated two-factor model representing physical and mental health [21].

Given that more and more researchers are expected to use PROMIS measures for patient-reported outcome assessments, it seems crucial to enable comparisons of respective results to other (e.g. older) studies using the SF-36. The aims of this study were therefore to establish easy-to-apply algorithms to reliably convert PROMIS-29 data to SF-36 summary component scores, and to validate these algorithms in new data not used for parameter estimation, in patients with cardiovascular disorders.

## Methods

### The Berlin long-term observation of vascular events (BeLOVE) study

The BeLOVE study is an ongoing long-term prospective observational cohort study of patients at very high risk for future cardiovascular events [24]. To meet inclusion criteria, patients must be at least 18 years of age and either recently hospitalized for an acute cardiovascular event (CVE) (acute coronary syndrome, acute heart failure, acute cerebrovascular disorder, and acute kidney injury) or at very high risk chronic cardiovascular conditions without event in the past 12 months. Pregnancy or breastfeeding, lack of health insurance, and life expectancy of  $\leq 6$  months due to a non-cardiovascular cause, active cancer, or a history of organ transplantation at the time of inclusion were defined as exclusion criteria. Moreover, patients unable to provide written informed consent are not considered for participation. Recruitment started in 2017 at the clinical campuses of the Charité - Universitätsmedizin Berlin and is ongoing.

The aim of BeLOVE is to improve prediction and understanding of disease progression and outcomes in patients with a very high risk of cardiovascular events, both in the acute and chronic phase, to ultimately improve and further personalize disease management. Assessments include comprehensive deep clinical and molecular phenotyping as well as ascertainment of clinical outcomes, e.g. major adverse CVEs, at predefined visits for up to 10 years.

In addition to clinical parameters, patient-reported outcome measures, including the PROMIS-29 profile and the SF-36, are administered at several assessment points of the BeLOVE study. Study data are collected and managed using REDCap [25, 26]. The present study utilized data from patients who were recruited during the first study phase of BeLOVE between July 2017 and

December 2020 and had participated in the PROM collection part of the study.

## Measures

### SF-36 physical (PCS) and mental (MCS) component summary scores

The SF-36 consists of eight domains: physical functioning (PF, 10 items), role function physical (RP, 4 items), bodily pain (BP, 2 items), general health (GH, 5 items), vitality (VT, 4 items), social functioning (SF, 2 items), role function emotional (RE, 3 items), and mental health (MH, 5 items) [14]. Scores of each domain can be transformed to a 0-100 scale.

The domain scores can be aggregated to physical (PCS) and mental (MCS) component summary scores; higher scores are representing better physical or mental health [16]. A norm-based T-score metric is used for scoring both the PCS and the MCS, with a mean of 50 and a standard deviation (SD) of 10 in the U.S. general population [16]. SF-36 PCS and MCS scores were originally derived using an orthogonal factor model, 'forcing' physical and mental components to be uncorrelated [16]. Since this original approach leads to potential problems with interpretation of results [17–19], modified scoring algorithms for correlated, i.e., oblique, SF-36 component summary scores (PCS<sub>c</sub> and MCS<sub>c</sub>), have been developed [20]. In the present study, we used component summary scores from both the orthogonal and the oblique factor solution, based on the German version of the standard SF-36 instrument with original recall periods ('the past 4 weeks' for most items) [16, 27].

### The PROMIS-29 v2.0 profile

The PROMIS initiative, which was funded by the U.S. National Institutes of Health, developed item banks for many physical and psychosocial self-reported health domains [23]. All items of a given item bank are calibrated to a unidimensional T-score metric with a general population mean of 50 and a SD of 10, using IRT modeling [28]. A main advantage of IRT-calibrated item banks is that any item subset (e.g., short form) can be used to yield T-scores on a standardized scale [29, 30]. The PROMIS-29 v2.0 profile consists of 4-item short forms of seven HRQL domains (pain interference, fatigue, depression, anxiety, sleep disturbance, physical function, and ability to participate in social roles) and an additional single item measuring pain intensity on a 0–10 numeric rating scale [10, 21].

Analogous to the SF-36, the domains of the PROMIS-29 can be aggregated to physical and mental health summary scores, which are based on a correlated factor solution [21]; higher scores indicate better health. Many PROMIS measures have been translated into other

languages, including German [31–33]. This study used the German version of the PROMIS-29 v2.0 profile [31].

### Study samples

Within the BeLOVE study, both the PROMIS-29 and the SF-36 were performed during the deep phenotyping visits ~90 days (visit 3, V3) and two years (visit 6, V6) after the qualifying CVE for the acute disease entities or following study inclusion in the chronic CV arm in the BeLOVE Unit of the Berlin Institute of Health at Charité - Universitätsmedizin Berlin. Because most SF-36 and PROMIS-29 domains consists of few items and to ensure stable estimates, data from participants who did not answer all items of both measures were excluded for further analysis; this approach has been applied before [20]. In the present study, we used V3 data to establish algorithms to predict SF-36 summary scores from PROMIS-29 ('calibration sample'), while V6 data were used to validate these algorithms ('validation sample').

### Sample size considerations

With regard to the calibration sample, a minimum sample size of 509 was calculated to be sufficient for detecting a small effect ( $f^2 > 0.02$ ) in a linear regression model with eight predictors (power=0.80, significance level=0.05). A minimum sample size of 180 in the validation sample was calculated for detecting small effect sizes, defined as a standardized mean difference (SMD) of  $> 0.20$  (power=0.80, significance level=0.05).

### Statistical analysis

Based on data from the calibration sample, we fitted four separate linear regression models each for predicting SF-36 physical and SF-36 mental component scores [34]. These regression models differed by both the dependent variables (uncorrelated versus correlated SF-36 component summary scores) and the predictors (PROMIS-29 domain scores versus PROMIS-29 physical/mental summary scores).

For each model, assumptions of (multiple) linear regression analysis were checked [34]. We inspected partial regression plots to rule out non-linear relationships between dependent and independent variables. To identify outliers potentially biasing the regression model, we calculated Cook's distance values (cut-off  $< 1$ ). To test the assumption of independent residuals, we used the Durbin-Watson statistic [35], which should be close to a value of 2. Homoscedasticity was checked graphically [36]. Variance Inflation Factors were calculated to rule out multicollinearity in those models with multiple predictors (cut-off  $< 10$ ).

We then applied the established regression coefficients to predict SF-36 physical and mental component summary scores from PROMIS-29 data in the validation

sample. Pearson correlation coefficients ( $r$ ) were calculated to determine the association between empirical (i.e., 'observed') and predicted SF-36 summary scores. For calculating SMDs for paired samples, we utilized a pragmatic approach as described by Cumming (2012), which is appropriate for determining within-group effect sizes [37]. Specifically, we used the formula:  $SMD = \text{mean difference between both measurements} / \text{averaged standard deviation}$  [37]. We considered SMD values of 0.2, 0.5, and 0.8 as small, medium, and large effects, respectively; values below 0.2 were considered negligible [38]. Mean absolute errors (mae), and root mean square errors (rmse) were used to compare the agreement between empirical and predicted scores across the different regression models [39, 40]. Smaller rmse and mae values indicate better agreement between empirical and predicted scores. Typically, the rmse is larger than the mae due to its sensitivity to larger errors.

For statistical analyses, R version 4.2.1 and the R packages 'Metrics', 'effize' and 'pwr' were used [40–43].

## Results

### Sample characteristics

Data from  $n=662$  and  $n=259$  patients with complete SF-36 and PROMIS 29 responses were used for calibration and validation analyses, respectively. Detailed sample characteristics with respect to age, gender, diagnosis that led to study inclusion, as well as SF-36 and PROMIS-29 scores are presented in Table 1.

SF-36 component summary scores and PROMIS-29 health summary scores indicated slightly better physical and mental health in the validation sample than in the calibration sample. However, these differences were less than 2 T-scores on a scale with a SD of about 10, corresponding to negligible effect sizes.

### Calibration of regression coefficients

Assumptions of (multiple) linear regression analysis were met for all fitted models. Table 2 summarized the regression results for both the uncorrelated (i.e., original) and correlated SF-36 component summary scores as outcomes, and with different predictors (i.e., PROMIS-29 domain score models versus PROMIS-29 summary score models).

### Uncorrelated (original) SF-36 component summary scores

For predicting the SF-36 PCS, adjusted  $R^2$  values were high for both the PROMIS-29 domain score model (76%) and the model with the PROMIS-29 physical summary score as single predictor (71%). In the PROMIS-29 domain score model, the strongest predictors of the SF-36 PCS were physical function, pain intensity, and pain interference.

**Table 1** Sample characteristics

	Calibration sample	Validation sample
<b>Sample size</b>	662	259
<b>Female; n (%)</b>	205 (31.0)	63 (24.3)
<b>Mean age (SD)</b>	62.8 (12.2)	62.6 (11.4)
<b>Median age (min.; max.)</b>	64 (20; 88)	63 (20; 84)
<b>Study arm; n (%)</b>		
Stroke	295 (44.6)	119 (45.9)
Diabetes	168 (25.4)	51 (19.7)
Myocardial infarction	145 (21.9)	74 (28.6)
Heart failure	29 (4.4)	12 (4.6)
Acute renal failure	25 (3.8)	3 (1.1)
<b>SF-36 Scores; mean (SD)</b>		
PF	68.8 (27.0)	71.9 (27.4)
RP	56.0 (41.9)	68.2 (40.6)
BP	68.8 (27.3)	68.0 (25.2)
GH	56.8 (19.3)	59.7 (20.3)
VT	56.8 (20.1)	59.6 (20.9)
SF	76.6 (23.2)	79.9 (24.0)
RE	70.0 (40.5)	73.6 (39.7)
MH	72.9 (17.5)	75.4 (17.4)
<b>PCS</b>	<b>42.5 (11.2)</b>	<b>44.1 (11.5)</b>
<b>MCS</b>	<b>49.6 (10.7)</b>	<b>50.7 (11.1)</b>
<b>PCS<sub>c</sub></b>	<b>43.5 (10.8)</b>	<b>45.4 (10.8)</b>
<b>MCS<sub>c</sub></b>	<b>46.9 (9.7)</b>	<b>48.5 (9.8)</b>
<b>PROMIS-29 Scores; mean (SD)</b>		
Physical function	47.2 (8.3)	48.6 (8.3)
Fatigue	48.3 (9.3)	47.2 (9.6)
Depression	49.8 (7.9)	48.6 (7.7)
Anxiety	50.5 (7.7)	49.4 (7.7)
Sleep disturbance	49.2 (8.5)	48.5 (7.4)
Pain interference	51.7 (9.3)	51.4 (8.6)
Ability to participate	51.8 (9.0)	52.5 (9.5)
Pain intensity	51.6 (10.8)	51.6 (10.1)
<b>Physical summary</b>	<b>47.6 (8.6)</b>	<b>48.9 (8.6)</b>
<b>Mental summary</b>	<b>50.9 (8.1)</b>	<b>51.9 (8.1)</b>

Abbreviations: max., Maximum; min., Minimum; n, number; SD, standard deviation

When using the PROMIS-29 to predict the SF-36 MCS, considerably less variation could be explained, compared to predicting the SF-36 PCS. In the domain score model, the adjusted  $R^2$  value was 57%, with depression and anxiety being the strongest predictors of the SF-36 MCS. Using the PROMIS-29 mental summary score as single predictor, only 40% of variation in the SF-36 MCS could be explained.

#### Correlated SF-36 component summary scores

When using the PROMIS-29 physical health summary score for predicting the SF-36 PCS<sub>c</sub>, the adjusted  $R^2$  value was comparably high to the model with the uncorrelated PCS as outcome (70%). In the multiple regression model with individual PROMIS-29 domain scores as predictors, even more variation of the PCS<sub>c</sub> could be explained

(adjusted  $R^2=81\%$ ), with physical function and pain intensity being the strongest predictors.

For predicting the SF-36 MCS<sub>c</sub>, the adjusted  $R^2$  value was also high for both the PROMIS-29 domain score model (74%) and the model with the PROMIS-29 mental health summary score as single predictor (71%). In the PROMIS-29 domain score model, the strongest single predictor of the SF-36 MCS<sub>c</sub> was depression.

#### Validation of regression models

##### Uncorrelated (original) SF-36 component summary scores

Results of applying the previously established regression coefficients to predict original SF-36 PCS and MCS scores from PROMIS-29 data in the validation sample are presented in Table 3. Pearson correlation coefficients between empirical and predicted SF-36 PCS scores were high, with  $r=0.87$  for the PROMIS-29 domain score

**Table 2** Regression results based on the calibration sample

Regression models with different predictors (independent variables)	Dependent variables			
	Uncorrelated summary scores		Correlated summary scores	
	PCS	MCS	PCS <sub>c</sub>	MCS <sub>c</sub>
<b>PROMIS-29 domain scores</b>				
Adjusted R-squared	76.1%	56.8%	80.6%	74.0%
Intercept	17.947*	104.129*	44.890*	82.026*
Slopes				
Physical function <sup>1</sup>	0.649*	-0.266*	0.415*	0.046
Fatigue <sup>1</sup>	-0.067	-0.198*	-0.147*	-0.210*
Depression <sup>1</sup>	0.121*	-0.476*	-0.077	-0.330*
Anxiety <sup>1</sup>	0.093*	-0.351*	-0.062	-0.233*
Sleep disturbance <sup>1</sup>	-0.011	-0.118*	-0.060*	-0.104*
Pain interference <sup>1</sup>	-0.369*	0.082	-0.277*	-0.073
Ability to participate <sup>1</sup>	0.145*	0.204*	0.221*	0.202*
Pain intensity <sup>2</sup>	-0.514*	-0.030	-0.428*	-0.177
<b>PROMIS-29 physical summary</b>				
Adjusted R-squared	70.5%	-	69.8%	-
Intercept	-9.563*	-	-6.100*	-
Slope	1.094*	-	1.041*	-
<b>PROMIS-29 mental summary</b>				
Adjusted R-squared	-	40.3%	-	71.3%
Intercept	-	6.815*	-	-4.663*
Slope	-	0.840*	-	1.013*

Abbreviations: MCS, uncorrelated Short Form-36 mental component score; MCS<sub>c</sub>, correlated Short Form-36 mental component score; PCS, uncorrelated Short Form-36 physical component score; PCS<sub>c</sub>, correlated Short Form-36 physical component score; PROMIS-29, Patient-Reported Outcomes Measurement Information 29-item profile measure v2.0

<sup>1</sup> T-scores with a general population mean = 50 (standard deviation = 10) based on 4-item short forms

<sup>2</sup> Single 0–10 numeric pain rating item

\* Predictor is statistically significant ( $p < 0.05$ )

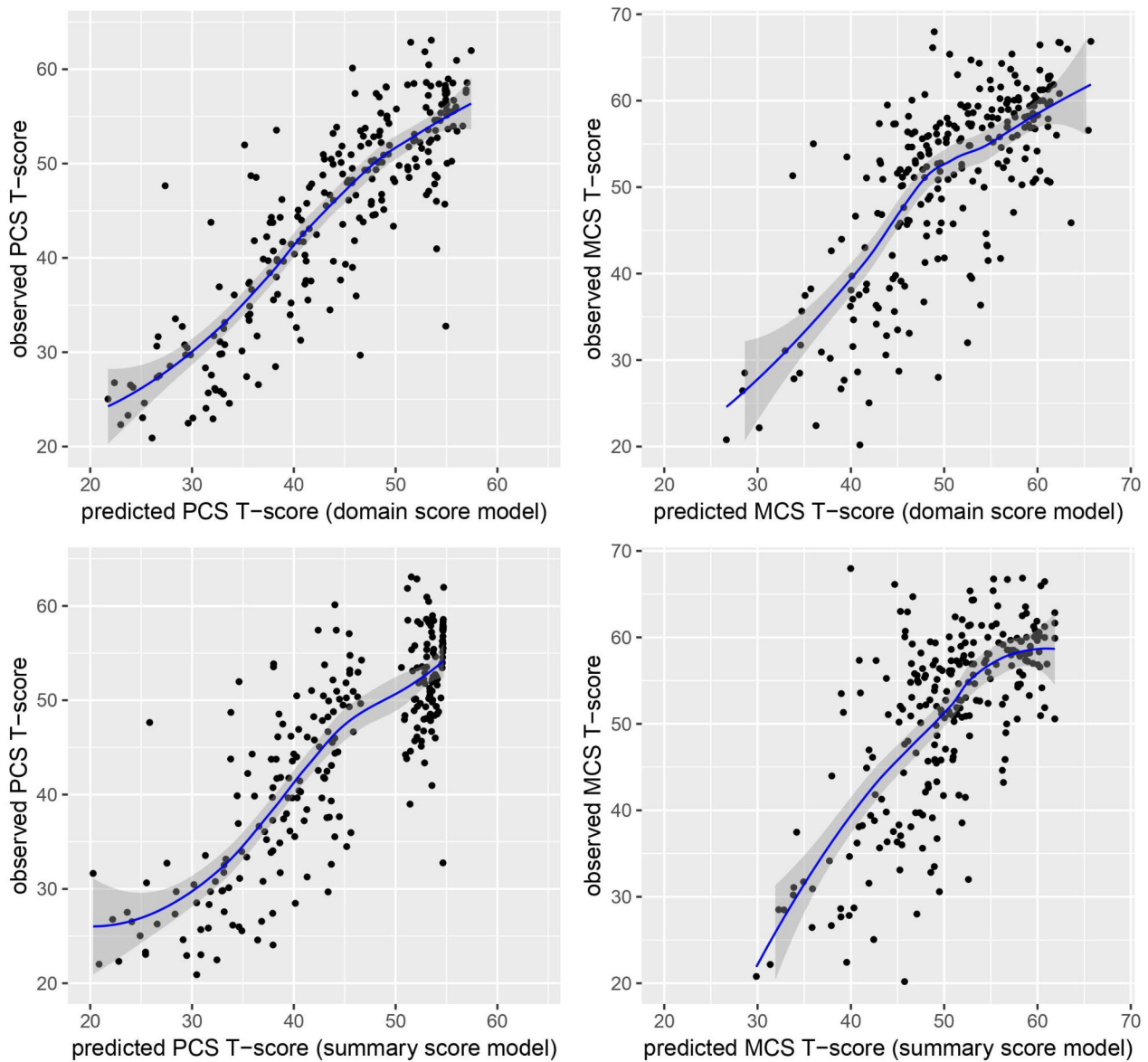
**Table 3** Validation results for uncorrelated (original) SF-36 component summary scores

Statistics	PCS		MCS	
	Empirical	Predicted	Empirical	Predicted
<b>PROMIS-29 domain score model</b>				
Mean (SD)	44.1 (11.5)	43.5 (9.5)	50.7 (11.1)	50.5 (8.1)
Pearson correlation		0.87		0.75
SMD [95% CI]		-0.06 [-0.12, 0.01]		-0.02 [-0.11, 0.07]
rmse		5.68		7.44
mae		4.33		5.76
<b>PROMIS-29 summary score model</b>				
Mean (SD)	44.1 (11.5)	44.0 (9.4)	50.7 (11.1)	50.4 (6.8)
Pearson correlation		0.83		0.68
SMD [95% CI]		-0.02 [-0.09, 0.05]		-0.03 [-0.13, 0.07]
rmse		6.44		8.20
mae		4.97		6.23

Abbreviations: CI, confidence interval; MCS, uncorrelated Short Form-36 mental component score; mae, mean absolute error; PCS, uncorrelated Short Form-36 physical component score; PROMIS-29, Patient-Reported Outcomes Measurement Information 29-item profile measure v2.0; rmse, root mean square error; SD, standard deviation; SMD, standardized mean difference

model and  $r=0.83$  for the PROMIS-29 summary score model. With regard to predicting SF-36 MCS scores, the association between empirical and predicted scores were lower, with  $r=0.75$  for the PROMIS-29 domain score model and  $r=0.68$  for the PROMIS-29 summary

score model. Related scatter plots are presented in Fig. 1, showing that predicted scores appear to be generally less biased in the domain score model as compared to the summary score models. Predicted PCS scores based on PROMIS-29 summary scores indicated ceiling effects.



**Fig. 1** Scatter plots showing the associations between predicted (x-axis) and observed (y-axis) SF-36 component summary scores (uncorrelated model)

On the group level, empirical and predicted SF-36 summary scores differed only negligibly, which is true for the PCS and the MCS, and for both the PROMIS-29 domain score model and the PROMIS-29 summary score model (SMDs between -0.06 and -0.02). However, the agreement between empirical and predicted summary scores, as assessed with the rmse and the mae, was better for predicting the PCS than the MCS.

**Correlated SF-36 component summary scores**

Measurement characteristics after applying the established regression coefficients to predict correlated SF-36 PCS<sub>c</sub> and MCS<sub>c</sub> scores from PROMIS-29 data in the validation sample are presented in Table 4. Pearson

correlation coefficients between empirical and predicted SF-36 component summary scores were generally high, with  $r \geq 0.85$  for all regression models. Scatter plots again indicated ceiling effects, when PROMIS-29 summary scores were used to predict PCS<sub>c</sub> scores (see Fig. 2).

The differences between empirical and predicted SF-36 summary score on the group level were negligible in each model (SMDs between -0.07 and -0.03). In contrast to the regression models with uncorrelated SF-36 PCS and MCS scores, the agreement between empirical and predicted summary scores, as assessed with the rmse and the mae, was comparably high between models with PCS<sub>c</sub> and MCS<sub>c</sub> scores as outcomes.

**Table 4** Validation results for correlated SF-36 component summary scores

Statistics	PCS <sub>c</sub>		MCS <sub>c</sub>	
	Empirical	Predicted	Empirical	Predicted
<b>PROMIS-29 domain score model</b>				
Mean (SD)	45.4 (10.8)	44.6 (9.4)	48.5 (9.8)	48.1 (8.3)
Pearson correlation		0.90		0.86
SMD [95% CI]		-0.07 [-0.13, -0.02]		-0.03 [-0.12, 0.01]
rmse		4.82		5.03
mae		3.74		3.99
<b>PROMIS-29 summary score model</b>				
Mean (SD)	45.4 (10.8)	44.8 (8.9)	48.5 (9.8)	47.9 (8.2)
Pearson correlation		0.85		0.87
SMD [95% CI]		-0.05 [-0.12, 0.01]		-0.06 [-0.10, 0.03]
rmse		5.68		4.90
mae		4.53		3.90

*Abbreviations:* CI, confidence interval; MCS<sub>c</sub>, correlated Short Form-36 mental component score; mae, mean absolute error; PCS<sub>c</sub>, correlated Short Form-36 physical component score; PROMIS-29, Patient-Reported Outcomes Measurement Information 29-item profile measure v2.0; rmse, root mean square error; SD, standard deviation; SMD standardized mean difference

## Discussion

Based on a sample from a large cardiovascular cohort, we established and validated regression coefficients that can be used to convert PROMIS-29 data to SF-36 physical and mental component summary scores. Satisfactory model fit was confirmed by applying the regression coefficients to new data from a subsequent observation of the same cohort, supporting validity of the established linear regression models in patients with cardiovascular diseases.

We found that using all eight PROMIS-29 domains to predict SF-36 component summary scores tended to produce slightly better results than using PROMIS-29 health summary scores as single predictors. Thus, if PROMIS-29 domain scores are available, we recommend using them directly to estimate SF-36 scores and avoid the intermediate step of calculating PROMIS-29 health summary scores. However, if only summary scores are available, they can also be used as reliable predictors.

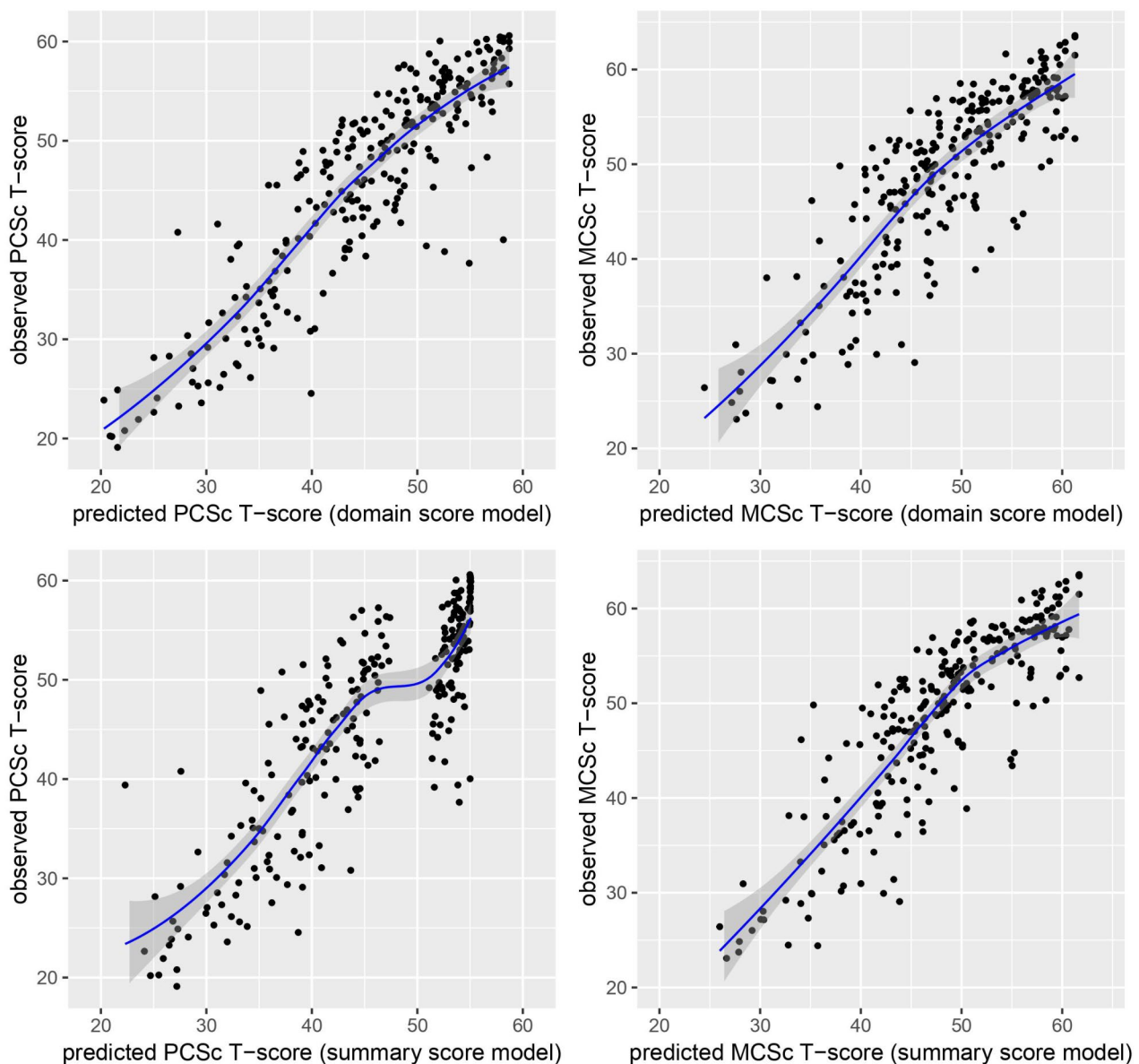
Regarding the prediction of SF-36 physical component summary scores, we found comparable results for either the original (PCS) or the correlated factor model (PCS<sub>c</sub>). The predictive power of PROMIS-29 scores was high for both the PCS and the PCS<sub>c</sub>. In the PROMIS-29 domain score models, the pattern of individual predictors was very similar, with physical function and pain being the strongest predictors for PCS and PCS<sub>c</sub> scores.

In contrast, for predicting SF-36 mental component summary scores using the PROMIS-29, considerably better predictive power and less biased predictions were found under the correlated (MCS<sub>c</sub>) than under the original factor model (MCS). A potential reason for this is that SF-36 component summary scores under an orthogonal (i.e., uncorrelated) factor model might be biased, which has been discussed before [17–20]. For example, we found that low PROMIS physical function scores were

significantly associated with higher MCS scores, which does not seem plausible. For predicting MCS<sub>c</sub> scores based on individual PROMIS-29 domains, this was not the case. A further reason for the particularly high association between PROMIS-29 mental summary scores and SF-36 MCS<sub>c</sub> scores might be that PROMIS health summary scores were also established under a correlated factor model [21].

This study has limitations. First, the representativeness of our sample might be affected by self-selection since patients in the cohort refused to participate in the PROM part of the study. Unfortunately, we do not know how the sample used for analysis differs from the full cohort with regard to self-reported health. However, both PROMIS-29 and SF-36 scores covered the full range of possible T-scores, indicating that the established regression coefficients are reliable for subpopulations with differing health status. A second limitation is that our study is based on data from German patients with cardiovascular diseases. It remains to be investigated whether our results can be generalized to non-German populations and to populations with non-cardiovascular disorders. Third, PROMIS-29 physical health summary scores showed ceiling effects, supporting the recommendation to preferably use individual PROMIS-29 domain scores for predicting SF-36 summary component scores. A fourth limitation is that many alternative methods for mapping scales are available [8, 44–46], which we did not employ. We chose linear regression for its simplicity, making our algorithms accessible to a broad audience within a well-known framework. Moreover, we consistently included all PROMIS-29 domains in our multiple regression models (and avoided stepwise selection of predictors [7]) to maintain consistency with the algorithms used for calculating PROMIS and SF-36 component scores [16, 20, 21]. In this context, we also experimented with polynomial





**Fig. 2** Scatter plots of the validation data showing the associations between predicted (x-axis) and observed (y-axis) SF-36 component summary scores (correlated model)

regression coefficients [7], which did not yield notable improvements in predictive power (data not shown) compared to the linear models. Furthermore, we did not employ bidirectional mapping methods, such as equipercentile equating or IRT-based linking methods [8, 44, 45], as our aim was to use PROMIS data on domain-level to predict SF-36 composite scores (which was empirically supported by yielding superior results compared to using PROMIS composite scores as predictors).

Despite these limitations, the results of this study are of practical importance for future research, particularly for measuring overall self-reported HRQL in cardiovascular populations. Measures of overall HRQL might be

advantageous over the use of individual health domains in some research contexts [21]. Moreover, the number of statistical comparisons, e.g., in studies of treatment efficacy, can be reduced by using health component scores [13, 16]. The SF-36 is probably the most frequently used generic measure of self-reported HRQL. However, the PROMIS-29, which is a newer measure based on modern test theory methods, has been discussed to be even more appropriate for assessing HRQL component scores than the SF-36 [47]. A particular advantage of the flexible PROMIS approach is that any item set of a given domain can be used to establish domain-related T-scores [23]. Consequently, PROMIS-29 health summary scores

can also be produced based on other items than those included in the PROMIS-29 profile measure, as long as T-score estimates for each of its eight domains can be provided.

In this context, by showing high associations with SF-36 component summary scores, the findings of our study confirmed construct validity of PROMIS-29 health summary scores. In view of these advantages, we expect PROMIS to be increasingly used for HRQL assessments, highlighting the usefulness of valid comparisons to research findings based on the SF-36.

## Conclusions

In sum, this study will help facilitating comparison and pooling of findings from the SF-36 and the PROMIS-29 profile, two of the most frequently used generic measures of self-reported HRQL. We hope that our study will encourage other researchers to replicate our models in other patient populations.

## Abbreviations

BeLOVE	Berlin Long-term Observation of Vascular Events
BP	Bodily Pain
CVE	Cardiovascular Event
GH	General Health
HRQL	Health-Related Quality of Life
IRT	Item Response Theory
mae	Mean Absolute Error
MCS <sub>c</sub>	Correlated Mental Component Summary Scores
MCS	Original Mental Component Summary Scores
MH	Mental Health
PCS <sub>c</sub>	Correlated Physical Component Summary Scores
PCS	Original Physical Component Summary Scores
PF	Physical Functioning
PROM	Patient-Reported Outcome Measure
PROMIS	Patient-Reported Outcomes Measurement Information System
RE	Role Function Emotional
rmse	Root Mean Square Error
RP	Role Function Physical
SD	Standard Deviation
SF	Social Functioning
SMD	Standardized Mean Difference
VT	Vitality

## Acknowledgements

We would like to thank the BIH for financial, operational and infrastructural support of this interdisciplinary cohort. Additionally, the authors wish to thank all participants in the BeLOVE pilot and main phase.

## Author contributions

G.L. contributed to study conception, analyzed and interpreted the data, and drafted the manuscript. F.H.F. contributed to analysis and interpretation of the data. C.N.M., M.R., M.Rö., A.B., and S.S. contributed to study conception and design, and interpretation of the data. M.A., L.H.B., K.U.E., M.E., F.E., H.G., U.G., A.H., N.H., U.L., D.L., K.M., D.N.M., C.H.N., K.M.S.O., T.P., J.S.M., J.S., J.E.W., and M.W. made substantial contributions to study conception and design and acquisition of data. J.K.H., S.K.P., S.R., O.S., K.S., and I.R.N. contributed to the processing and provision of the data. All authors were involved in revising the article critically for important intellectual content, and gave final approval of the version to be submitted.

## Funding

The costs related to the BeLOVE study are funded by the Berlin Institute of Health at Charité-Universitätsmedizin Berlin. Open Access funding enabled and organized by Projekt DEAL.

## Data availability

The BeLOVE consortium aims to ensure that the collected data and sample material will be used for the greatest possible benefit to health-related research, in particular cardiovascular research. Researchers interested in the data of BeLOVE may apply for data access through our use and access committee, as long as one member of the project team is part of the BIH research community to support the research process. The use and access committee evaluates the merits and technical feasibility of the project proposal and assesses potential overlap with ongoing projects and analyses. Data transfer will be performed according to established General Data Protection Regulation (GDPR) data sharing guidelines.

## Declarations

### Ethics approval and consent to participate

Approval was granted by the Ethics Committee of Charité – Universitätsmedizin Berlin (Date of vote: 2017-07-06; Ethic committee number EA1/066/17). All study participants are informed comprehensively and with sufficient time for them to reflect on the nature and scope of the study. Written informed consent is obtained from all participants prior to all study-related procedures. Informed consent to participate includes permission to analyse data and samples and to publish all results, at least by the BeLOVE Group. BeLOVE was registered at the approved WHO primary register: German Clinical Trials Register: <https://drks.de/search/de/trial/DRKS00016852>; WHO International Clinical Registry Platform: <http://apps.who.int/trialsearch/Trial2.aspx?TrialID=DRKS00016852>.

### Competing interests

Frank Edelmann (FE) reports grants from German Research Foundation (DFG), grants from German Ministry of Education and Research, grants from the German Herta Foundation; during the conduct of the study; personal fees and non-financial support from Novartis, grants and personal fees from Boehringer Ingelheim, personal fees from CVRx, Pfizer, Medtronic, grants and personal fees from Servier, personal fees from MSD, personal fees from Merck & Co., grants from AstraZeneca, personal fees from Bayer, personal fees from Resmed, personal fees from Berlin Chemie, grants from Thermo Fischer, personal fees from Vifor Pharma, personal fees from PharmaCosmos outside the submitted work. Matthias Endres (ME) reports grants from Bayer and fees paid to the Charité from Abbot, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, BMS, Daiichi Sankyo, Amgen, Sanofi, , Novartis, Pfizer, all outside the submitted work. ME received funding from DFG under Germany's Excellence Strategy – EXC-2049 – 390688087, Collaborative Research Center ReTune TRR 295- 424778381, BMBF, DZNE, DZHK, EU, Corona Foundation, and Fondation Leducq. Holger Gerhardt (HG) reports grants from the DFG, the Leducq Foundation, the BMBF and the German Center for Cardiovascular Research (DZHK) during the conduct of the study, outside of the submitted work. Ulf Landmesser (UL) reports research funding from German Cardiovascular Research Network (DZHK); Fondation Leducq; research grants from Novartis, Bayer and Amgen. Knut Mai (KM) declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported. Gregor Liegel (GL) received funding for research from the German Research Foundation (DFG). Dominik N. Müller (DNM) received funding for research from Bayer Healthcare, Deutsche Forschungsgemeinschaft and from BMBF. Christian H Nolte (CHN) received research grants from German Ministry of Research and Education, German Center for neurodegenerative Diseases (DZNE), German Center for cardiovascular Research (DZHK), and speaker and/or consultation fees from Boehringer Ingelheim, Bristol-Myers Squibb, Pfizer Pharma, Abbott, Novartis, Daichii-Sankyo and Alexion all outside the submitted work. Tobias Pischon (TP) received grants from the Federal Ministry of Education and Research (BMBF), the Federal Ministry of Food and Agriculture (BMEL), the Federal Ministry for Economic Affairs and Energy (BMWi), the Deutsche Forschungsgemeinschaft (DFG), Deutsche Herzstiftung, German Academic Exchange Service (DAAD). Kai M. Schmidt-Ott (KMSO) reports receiving research funding from Deutsche Forschungsgemeinschaft (CRC 1365, GRK 2318, RU 2841), Urological Research Foundation Berlin, ERA PerMed (OnAKI-ICI), COVID-19-Forschungsnetzwerk Niedersachsen (EXTINCT post COVID), Niedersächsisches Ministerium für Wissenschaft und Kultur (zukunft.Niedersachsen), FAST BioMedical, Quark Pharmaceuticals, REATA, Boehringer Ingelheim, AstraZeneca, Sanofi, Alentis, Roche, and Alexion; having received consultancy fees from BioPorto Diagnostics and Abionyx; having participated in advisory boards with Bayer, Stadapharm, and

Boehringer Ingelheim; having received license revenue related to the use of a neutrophil gelatinase-associated lipocalin assay via Columbia University; and being an editorial board member of *Kidney International* and *Kidney International Reports*; each outside the submitted work. Jeanette Schulz-Menger (JSM) reports grants from Bayer Healthcare, non-financial support from Siemens healthineers, non-financial support from Circle cardiovascular, non-financial support from Medis, outside the submitted work, and Bayer Healthcare, Advisor. Furthermore, funding for research from the EU, DZHK, Deutsche Herzstiftung. Jochen Spranger (JS) received funding for research from Deutsche Forschungsgemeinschaft and from BMBF. Martin Witznath (MW) received funding for research from Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, Deutsche Gesellschaft für Pneumologie, European Respiratory Society, Marie Curie Foundation, Else Kröner Fresenius Foundation, Capnetz Foundation, International Max Planck Research School, Actelion, Bayer Health Care, Biotest, Boehringer Ingelheim, Noxxon, Pantherna, Quark Pharma, Vaxxilon, and for lectures and advisory from Actelion, Aptarion, Astra Zeneca, Bayer Health Care, Berlin Chemie, Biotest, Boehringer Ingelheim, Chiesi, Glaxo Smith Kline, Novartis, Noxxon, Pantherna, Teva and Vaxxilon. All remaining authors do not report potential conflicts of interest.

#### Author details

<sup>1</sup>Department of Psychosomatic Medicine, Center for Patient-Centered Outcomes Research, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany

<sup>2</sup>Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>3</sup>Center for Stroke Research Berlin, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>4</sup>Department of Neurology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>5</sup>Department of Nephrology and Medical Intensive Care, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>6</sup>Department of Internal Medicine and Cardiology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>7</sup>German Centre for Cardiovascular Research (DZHK), partner site Berlin, Berlin, Germany

<sup>8</sup>German Center for Neurodegenerative Diseases (DZNE), partner site Berlin, Berlin, Germany

<sup>9</sup>Excellence Cluster NeuroCure, Berlin, Germany

<sup>10</sup>Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

<sup>11</sup>Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>12</sup>ECRC, Department of Cardiology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>13</sup>Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>14</sup>Department of Cardiology, Angiology and Intensive Care Medicine, Deutsches Herzzentrum der Charité (DHZC), Berlin, Germany

<sup>15</sup>Department of Cardiology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>16</sup>Department of Endocrinology & Metabolism, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>17</sup>Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Charité-Center for Cardiovascular Research (CCR), Berlin, Germany

<sup>18</sup>Experimental and Clinical Research Center (ECRC), a cooperation of Charité - Universitätsmedizin Berlin and Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany

<sup>19</sup>Division of Pulmonary Inflammation, Department of Infectious Diseases and Respiratory Medicine, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany

<sup>20</sup>German Center for Lung Research (DZL), Giessen, Germany

<sup>21</sup>Molecular Epidemiology Research Group, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

<sup>22</sup>Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Biobank Technology Platform, Berlin, Germany

<sup>23</sup>German Center for Diabetes Research, München-Neuherberg, Germany

<sup>24</sup>Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Berlin, Germany

<sup>25</sup>Department of Nephrology and Hypertension, Hannover Medical School, Hannover, Germany

<sup>26</sup>Clinic for Psychosomatic Medicine and Psychotherapy, University of Duisburg-Essen, LVR-University Hospital Essen, Essen, Germany

<sup>27</sup>Centre for Translational Neuro- and Behavioral Sciences (C-TNBS), University of Duisburg-Essen, Essen, Germany

<sup>28</sup>Department of Medicine, Cardiology, Goethe University Hospital, Frankfurt, Germany

<sup>29</sup>German Center for Cardiovascular Research (DZHK) Partner Site RheinMain, Frankfurt, Germany

Received: 19 January 2024 / Accepted: 29 July 2024

Published online: 15 August 2024

#### References

1. Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *Eur J Health Econ.* 2018;19(1):123–52.
2. Burns DJP, Arora J, Okunade O, Beltrame JF, Bernardes-Pereira S, Crespo-Leiro MG, et al. International Consortium for Health Outcomes Measurement (ICHOM): standardized patient-centered outcomes measurement set for heart failure patients. *JACC Heart Fail.* 2020;8(3):212–22.
3. Hahn EA, Walsh MN, Allen LA, Lee CS, Denfeld QE, Teuteberg JJ, et al. Validity of patient-reported outcomes Measurement Information System Physical, Mental, and Social Health measures after Left Ventricular assist device implantation and implications for patient care. *Circ Cardiovasc Qual Outcomes.* 2023;16(2):e008690.
4. Marquis P, Caron M, Emery M-P, Scott JA, Arnould B, Acquardo C. The role of Health-Related Quality of Life Data in the drug approval processes in the US and Europe. *Pharm Med.* 2011;25(3):147–60.
5. Kornowski R. Patient-reported outcome measures in cardiovascular disease. *Eur Heart J - Qual Care Clin Outcomes.* 2021.
6. Puhan MA, Soesilo I, Guyatt GH, Schünemann HJ. Combining scores from different patient reported outcome measures in meta-analyses: when is it justified? *Health Qual Life Outcomes.* 2006;4:94.
7. Klapproth CP, van Bebber J, Sidey-Gibbons CJ, Valderas JM, Leplege A, Rose M, et al. Predicting EQ-5D-5L crosswalk from the PROMIS-29 profile for the United Kingdom, France, and Germany. *Health Qual Life Outcomes.* 2020;18(1):389.
8. Liegl G, Wahl I, Berghöfer A, Nolte S, Pieh C, Rose M, et al. Using Patient Health Questionnaire-9 item parameters of a common metric resulted in similar depression scores compared to independent item response theory model reestimation. *J Clin Epidemiol.* 2016;71:25–34.
9. Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Wiley; 2013.
10. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS® adult health profiles: efficient short-form measures of seven health domains. *Value Health.* 2019;22(5):537–44.
11. Ahmad FS, Kallen MA, Schifferdecker KE, Carluzzo KL, Yount SE, Gelow JM, et al. Development and initial validation of the PROMIS®-Plus-HF profile measure. *Circulation: Heart Fail.* 2019;12(6):e005751.
12. Reynolds MR, Walczak J, White SA, Cohen DJ, Wilber DJ. Improvements in symptoms and quality of life in patients with paroxysmal atrial fibrillation treated with radiofrequency catheter ablation versus antiarrhythmic drugs. *Circ Cardiovasc Qual Outcomes.* 2010;3(6):615–23.
13. Salisbury C, Montgomery AA, Hollinghurst S, Hopper C, Bishop A, Franchini A, et al. Effectiveness of PhysioDirect telephone assessment and advice

- services for patients with musculoskeletal problems: pragmatic randomised controlled trial. *BMJ: Br Med J*. 2013;346:f43.
14. Ware J, Snow K, Kosinski M, Gandek B. SF-36 health survey manual and interpretation guide. Boston, MA: Health Assessment Lab, New England Medical Center; 1993.
  15. White MK, Maher SM, Rizio AA, Bjorner JB. A meta-analytic review of measurement equivalence study findings of the SF-36® and SF-12® health surveys across electronic modes compared to paper administration. *Qual Life Res*. 2018;27(7):1757–67.
  16. Ware J, Kosinski M, Keller S. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Institute, New England Medical Center; 1994.
  17. Hann M, Reeves D. The SF-36 scales are not accurately summarised by independent physical and mental component scores. *Qual Life Res*. 2008;17(3):413–23.
  18. Simon GE, Revicki DA, Grothaus L, Vonkorff M. SF-36 summary scores: are physical and mental health truly distinct. *Med Care*. 1998;36(4):567–72.
  19. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? *Qual Life Res*. 2001;10(5):395–404.
  20. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V. 1. *Health Qual Life Outcomes*. 2007;5(1):1–8.
  21. Hays RD, Spritzer KL, Schalet BD, Cella D. PROMIS®-29 v2. 0 profile physical and mental health summary scores. *Qual Life Res*. 2018;27(7):1885–91.
  22. Elsmann EB, Roorde LD, Smidt N, de Vet HCW, Terwee CB. Measurement properties of the Dutch PROMIS-29 v2.1 profile in people with and without chronic conditions. *Qual Life Res*. 2022;31(12):3447–58.
  23. 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. *Med Care*. 2007;45(5 Suppl 1):S3.
  24. Weber JE, Ahmadi M, Boldt L-H, Eckardt K-U, Edelmann F, Gerhardt H, et al. Protocol of the Berlin Long-Term Observation of vascular events (BeLOVE): a prospective cohort study with deep phenotyping and long-term follow up of cardiovascular high-risk patients. *BMJ Open*. 2023;13(10):e076415.
  25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.
  26. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
  27. Bullinger M. German translation and psychometric testing of the SF-36 health survey: preliminary results from the IQOLA project. *Soc Sci Med*. 1995;41(10):1359–66.
  28. 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.
  29. Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*. 2007;16(1):133–41.
  30. Liegl G, Gandek B, Fischer HF, Bjorner JB, Ware JE, Rose M, et al. Varying the item format improved the range of measurement in patient-reported outcome measures assessing physical function. *Arthritis Res Therapy*. 2017;19(1):1–12.
  31. 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(4):999–1014.
  32. 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. *Int J Methods Psychiatr Res*. 2017;26(4):e1530.
  33. Liegl G, Rose M, Knebel F, Stengel A, Buttgerit F, Obbarius A, et al. Using subdomain-specific item sets affected PROMIS physical function scores differently in cardiology and rheumatology patients. *J Clin Epidemiol*. 2020;127:151–60.
  34. Weisberg S. *Applied linear regression*. Wiley; 2005.
  35. Durbin J, Watson GS. Testing for serial correlation in least squares regression. I. *Biometrika*. 1950;37(3–4):409–28.
  36. Cook RD, Weisberg S. Diagnostics for heteroscedasticity in regression. *Biometrika*. 1983;70(1):1–10.
  37. Cumming G. *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis*. Routledge; 2013.
  38. Cohen J. *Statistical power analysis for the behavioral sciences*. routledge; 2013.
  39. Dewitt B, Jalal H, Hanmer J. Computing PROPr utility scores for PROMIS® profile instruments. *Value Health*. 2020;23(3):370–8.
  40. Hamner B, Frasco M, LeDell E. Package 'Metrics'. R Foundation for Statistical Computing; 2018.
  41. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
  42. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A et al. Package 'pwr'. R Package Version. 2018;1(2).
  43. Torchiano M, Torchiano MM. Package 'effsize'. Package Effsize. 2020.
  44. Blackwell CK, Tang X, Elliott AJ, Thomes T, Louwagie H, Gershon R, et al. Developing a common metric for depression across adulthood: linking PROMIS depression with the Edinburgh postnatal depression scale. *Psychol Assess*. 2021;33(7):610–8.
  45. ten Klooster PM, Oude Voshaar MAH, Gandek B, Rose M, Bjorner JB, Taal E, et al. Development and evaluation of a crosswalk between the SF-36 physical functioning scale and Health Assessment Questionnaire disability index in rheumatoid arthritis. *Health Qual Life Outcomes*. 2013;11(1):199.
  46. Keetharuth AD, Gray LA, McGrane E, Worboys H, Orozco-Leal G. Mapping Short Warwick and Edinburgh Mental Wellbeing Scale (SWEMWBS) to recovering quality of life (ReQoL) to estimate health utilities. *Health Qual Life Outcomes*. 2024;22(1):7.
  47. Alcantara J, Whetten A, Zabriskie C, Jones S. Exploratory factor analysis of PROMIS-29 V1.0, PROMIS Global Health and the RAND SF-36 from chiropractic responders attending care in a practice-based research network. *Health Qual Life Outcomes*. 2021;19(1):82.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.