

Measurement Properties of 2 Novel PROs, the Pompe Disease Symptom Scale and Pompe Disease Impact Scale, in the COMET Study

Mazen M. Dimachkie, MD, Priya S. Kishnani, MD, Cristina Ivanescu, PhD, Giulio Flore, MSc, Chad Gwaltney, PhD, Nadine A.M.E. van der Beek, MD, PhD, Alaa Hamed, MD, MPH, MBA, Kristina An Haack, MD, Laurence Pollissard, MSc, Eileen Baranowski, PharmD, MS, Susan E. Sparks, MD, PhD, and Pronabesh DasMahapatra, MD, MPH, for COMET Study Group

Correspondence

Dr. Pollissard
laurence.pollissard@sanofi.com

Neurology: Clinical Practice 2023;13:e200181. doi:10.1212/CPJ.0000000000200181

Abstract

Background and Objectives

The Pompe Disease Symptom Scale (PDSS) and Impact Scale (PDIS) were created to measure the severity of symptoms and functional limitations experienced by patients with late-onset Pompe disease (LOPD). The objectives of this analysis were to establish a scoring algorithm and to examine the reliability, validity, and responsiveness of the measures using data from the COMET clinical trial.

Methods

The COMET trial was a randomized, double-blind study comparing the efficacy and safety of avalglucosidase alfa and alglucosidase alfa in patients with LOPD aged 16–78 years at baseline. Adult participants (18 years or older) completed the PDSS and PDIS daily for 14 days at baseline and for 2 weeks before quarterly clinic visits for 1 year after randomization using an electronic diary. Data were pooled across treatment groups for the current analyses. Factor analysis and inter-item correlations were used to derive a scoring algorithm. Test-retest and internal consistency analyses examined the reliability of the measures. Correlations with criterion measures were used to evaluate validity and sensitivity to change. Anchor and distribution-based analyses were conducted to estimate thresholds for meaningful change.

Results

Five multi-item domain scores were derived from the PDSS (Shortness of Breath, Overall Fatigue, Fatigue/Pain, Upper Extremity Weakness, Pain) and 2 from the PDIS (Mood, Difficulty Performing Activities). Internal consistency (Cronbach $\alpha > 0.90$) and test-retest reliability (intraclass correlation > 0.60) of the scores were supported. Cross-sectional and longitudinal correlations with the criterion measures generally supported the validity of the scores ($r > 0.40$). Within-patient meaningful change estimates ranging from 1.0 to 1.5 points were generated for the PDSS and PDIS domain scores.

Discussion

The PDSS and PDIS are reliable and valid measures of LOPD symptoms and functional impacts. The measures can be used to evaluate burden of LOPD and effects of treatments in clinical trials, observational research, and clinical practice.

Trial Registration Information

ClinicalTrials.gov identifier: NCT02782741

University of Kansas Medical Center (MMD), Kansas City, KS; Duke University Medical Center (PSK), Durham, NC; IQVIA Netherland (CI, GF), Amsterdam-Zuidoost, North Holland Province, The Netherlands; Gwaltney Consulting (CG), Westerly, RI; Center for Lysosomal and Metabolic Diseases (NAMEB), Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; Sanofi (AH, EB, SES, PD), Cambridge, MA; and Sanofi (KAH, LP), Chilly-Mazarin, France.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at [Neurology.org/cp](https://www.neurology.org/cp).

The Article Processing Charge was funded by Sanofi.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Introduction

Late-onset Pompe disease (LOPD) is a rare disorder caused by a deficiency of acid alpha-glucosidase, leading to glycogen accumulation in lysosomes, especially within muscle cells.^{1,2} Although the various symptoms of LOPD (eg, muscle weakness and respiratory impairment) are described in the literature,² quantification of the frequency and impact of symptoms are not well-characterized. Therefore, Sanofi developed 2 patient-reported outcome (PRO) instruments: (1) a Pompe Disease Symptom Scale (PDSS), assessing breathing; feelings of tiredness, fatigue, or muscle weakness in different parts of the body; pain; and morning headache, and (2) a Pompe Disease Impact Scale (PDIS), assessing impacts of anxiety, feelings of worry and depression, and the ability and difficulties in performing certain activities of daily living (eg, walking, climbing stairs, rising from a sitting position, bending over, squatting down, exercising).

Qualitative development of the conceptual model has been reported in detail.³ In brief, findings from a detailed literature review were used to construct a preliminary conceptual model. Expert interviews were conducted with 3 LOPD-experienced clinicians, followed by 3 rounds of interviews with 13 participants with LOPD (4–5 participants/round), to discuss the initial model. Interview findings were used to identify novel and important key concepts and to finalize the conceptual model. The key areas in the conceptual model include signs and symptoms (respiratory, motor, and other) related to the disease process; direct impacts of symptoms on mobility and activity performance, eating, and activities of daily living; and general impacts on psychosocial experience and social and occupational participation.

The PDSS and PDIS were implemented as outcome measures in the Phase 3 COMET study (NCT02782741) comparing the efficacy and safety of avalglucosidase alfa with alglucosidase alfa during the study's double-blind phase. Treatment-naïve participants with a confirmed diagnosis of LOPD, aged 3 years and older, were enrolled into the COMET study.⁴ These tools are designed to measure changes in symptom severity and impacts of disease in patients with LOPD and also aid in the evaluation of the therapeutic efficacy of treatment throughout the LOPD disease spectrum in those 18 years and older.

We evaluated item performance, reliability, validity, and responsiveness of the PDSS and PDIS instruments using baseline and longitudinal data from patients with LOPD who were enrolled in the COMET study.

Methods

PDSS/PDIS Questionnaires

The PDSS and PDIS are self-administered questionnaires specifically designed to capture, respectively, symptoms and impacts of LOPD on individuals 18 years and older. Items

and scales for the questionnaires are shown in eFigure 1 (links.lww.com/CPJ/A445).

The PDSS is a 12-item numerical scale measuring the following LOPD symptoms: Breathing difficulties (items 1 and 2), Tiredness (item 3), Fatigue (item 4), Muscle weakness (items 5–9), Muscle ache (item 10), Pain (item 11), and Morning headache (item 12). Patients rate their symptoms using an 11-point scale from 'none' (0) to 'as bad as I can imagine' (10). Higher scores indicate greater severity of symptoms.

The PDIS is a 15-item numerical scale measuring the following LOPD impacts: Anxiety (item 1), Worry (item 2), Depression (item 3), Ability to walk (items 4 and 5), Ability to climb stairs (items 6 and 7), Rising from a sitting position (items 8 and 9), Ability to bend over (items 10 and 11), Ability to squat (items 12 and 13), and Ability to exercise (items 14 and 15). Mood items (Anxiety, Worry, and Depression) are rated using an 11-point scale from 'no impact' (0) to 'as bad as I can imagine' (10). For difficulty performing activities, there are 6 pairs of questions. The first of each pair asks whether the activity was performed in the past 24 hours (items 4, 6, 8, 10, 12, and 14), with patients answering 'no and not physically able' (0), 'no but physically able' (1), or 'yes' (2). Patients answering 'yes' are asked the second question in the pair, categorizing the degree of difficulty in performing the activity in the past 24 hours (items 7, 9, 11, 13, and 15, respectively). The second question is answered on a 5-point scale: 'not at all difficult' (0), 'a little difficult' (1), 'somewhat difficult' (2), 'very difficult' (3), and 'extremely difficult' (4). Patients responding with '0' or '1' on the first question do not answer the second question. Higher scores indicate greater impact on items.

The scoring approaches for the PDSS and PDIS are provided in full in eAppendix 1 and eAppendix 2 (links.lww.com/CPJ/A444), respectively.

Quantitative Analyses

During the COMET study,⁴ participants 18 years and older completed a 24-h recall e-diary at baseline and the data were averaged over 7 consecutive days (day –14 to day –8 and day –7 to day –1 before treatment). Thereafter, biweekly (14-day) scores were calculated as the average of daily responses from the protocol-specified visits (weeks 13, 25, 37, and 49 on treatment) and the 13 consecutive days before each visit. If the e-diary was unavailable (eg, device dysfunction or deficiency), a paper version was completed, and data were transferred to the database. If ≥ 4 daily scores of a 7-day period were missing, then the score was set to missing. Each COMET site explained the e-diary to participants. Alerts were sent to the participant's phone during the study to remind them of important study tasks.

Quantitative analyses were conducted on pooled data from the COMET study,⁴ that is, PDSS or PDIS data from the avalglucosidase alfa and alglucosidase alfa arms were pooled,

to evaluate item performance and psychometric properties of both questionnaires. All analyses were performed on the modified intent-to-treat (mITT) population,⁴ defined as all randomized patients who received at least one partial or total infusion of the study drug.

Distributional properties were assessed on item and domain characteristics. Summary statistics and the percentage of scale ranges were calculated for each symptom and impact item in the scales. The proportions of participants scoring at the most severe end of the scale (floor effect) and the least severe end of the scale (ceiling effect) for each item were also calculated. Floor and ceiling effects were considered small if $\leq 15\%$ of patients achieved the worst and best health state and serious if $> 15\%$ of patients achieved these states.⁵

Exploratory factor analysis (EFA)⁶ was conducted for the PDSS and PDIS separately, using an unweighted least square extraction method and promax rotation assuming that factors are correlated (oblique). The number of factors was determined using eigenvalues, screen plots, factor loading, simple structure, and clinical judgment. A total summary scale or set of scales were proposed considering the factor loadings and clinical and conceptual considerations. Item-to-item correlations and item-to-total correlations (corrected for overlap) were computed using Spearman rank correlation.

Internal consistency was evaluated using Cronbach alpha coefficients, with > 0.7 considered as supporting the reliability of internal consistency.⁷ Intraclass correlation coefficient (ICC) values of 0.50–0.90 are considered to represent moderate-to-good reliability and values > 0.90 excellent reliability.⁸ Test-retest reliability, an assessment of repeatable reliability (i.e., stability of an instrument over time), was assessed using a two-way mixed consistency model for the calculation of the ICCs (2,1) between 2 time points. Two test-retest analyses were conducted, one during screening (days -14 to -8 for the test and days -7 to -1 for the retest) and one between baseline (days -7 to -1) for the test and week 49 for the retest. The latter analysis was only conducted for participants who reported no change according to Patient Global Impression of Change (PGIC)⁹ items.

Construct validity assessed the ability of the PDSS and PDIS to measure the core constructs (i.e., the intended aspects of the disease model), by using data generated with other instruments and/or clinical assessments; both convergent validity and divergent validity were assessed. Convergent validity assessed the strength of correlation between instruments measuring the same concept, and divergent validity assessed the lack of correlation between instruments measuring different or weakly related concepts. Spearman correlation coefficients were calculated for the relationships of the PDSS and PDIS weekly scores and their EFA-derived scales at baseline (day -7 to day -1) with PRO scales/items from the following instruments: Short-Form 12 items (SF-

12)—Physical Component Summary (PCS) and Mental Component Summary (MCS) scales,¹⁰ EuroQol-5 dimensions 5 levels (EQ-5D-5L) Pain and Mobility scales,¹¹ and Rasch-built Pompe-specific activity (R-PAct) scale¹² for quality of life and disease-related symptoms and forced vital capacity (FVC) % predicted in the upright position^{13,14}; 6-minute walk test (6MWT) distance measured in meters¹⁵⁻¹⁷; and Quick Motor Function Test (QMFT)¹⁸ for outcome measures.

Sensitivity to change is the ability of an instrument to measure change in a state regardless of whether it is relevant or meaningful to the decision maker.¹⁹ This was assessed using an analysis of covariance adjusted for baseline scores to compare change from baseline in PDSS and PDIS at week 49 among those who improved in contrast with those who worsened or remained unchanged according to PGIC daily activities ratings at week 49. In addition, the magnitude of change in each change category using effect sizes (ESs, i.e., mean change from baseline divided by the SD at baseline) and standardized response means (SRMs, i.e., mean change from baseline divided by the SD of the change from baseline) was assessed. The responder groups as independent variables were defined as follows: ‘improved’: participants answering “a great deal better,” “moderately better,” or “somewhat better” and ‘worsened/no change’: participants answering “a great deal worse,” “moderately worse,” “somewhat worse,” or “no change.”

Within-patient meaningful change threshold. Both anchor and distribution-based methods were used to determine clinically meaningful improvement thresholds for the PDSS and PDIS. Anchor-based methods link PRO scores to known clinically relevant indicators or to the patient’s determined rating of change. Anchors must correlate at least moderately ($r > 0.30$) with the PROs²⁰ to be interpretable. The anchor-based approach evaluated meaningful change from baseline to week 49 using scores from PGIC items⁹ as anchors. These PGIC items consisted of Ability with Daily Activities (item 1), Disease-Related Symptoms (item 2), Ability to Breathe (item 3), and Mobility (item 4). Distribution-based methods, which allow change to be interpreted in the context of the variability of scores and reliability of the instrument, were also used to estimate the meaningful change threshold. Two distribution methods were used: one-half SD and standard error of measurement.

Standard Protocol Approvals, Registrations, and Patient Consents

This study used data from the COMET (NCT02782741) clinical trial. The COMET study protocol was reviewed and approved by appropriate ethics committees and/or institutional review boards and conducted in accordance with the Declaration of Helsinki and the International Council for Harmonisation guidelines for Good Clinical Practice. Research ethics and informed consent have previously been reported for the COMET study.⁴

Data Availability

Qualified researchers may request access to participant-level data and related study documents. Participant-level data will be anonymized, and study documents will be redacted to protect the privacy of study participants. Further details on data sharing criteria, eligible studies, and the process for requesting access of Sanofi can be found at vivli.org.

Results

Study Population

The mITT population included 100 participants from the COMET study, of whom 99 were 18 years or older and 1 was 16 years. The mean \pm SD age of COMET participants was 47.6 ± 14.2 years (range, 16–78 years); 52 were male (52%). Most participants were White (94 [94%]), and ethnicity was also unevenly split between 76 (76%) not Hispanic or Latino, 15 (15%) Hispanic or Latino, and 9 (9.0%) who did not report their ethnicity. Baseline clinical and PRO measures are provided in eTable 1 (links.lww.com/CPJ/A445). At baseline, there were up to 72 participants with valid responses to individual PDSS/PDIS questions. The 1 participant younger than 18 years did not complete the PDSS and PDIS.

The PDSS and PDIS instrument completion rate was around 63% of expected daily entries.

Score Distribution

The PDSS and PDIS score distribution of the averaged weekly values for the week leading up to the baseline is provided in Table 1. Ceiling effects indicating the lowest symptom severity were most notable for 4 PDSS items: item 1 (Breathing), item 2 (Breathing while lying down), item 8 (Muscle weakness hand), and item 12 (Morning headache) (eTable 2, links.lww.com/CPJ/A445). No floor effects indicating the highest symptom severity were observed at baseline.

Baseline ceiling effects were observed for 2 PDIS items—item 1 (Anxiety) and item 3 (Depression), respectively; floor effects indicating the highest symptom severity were noted for 3 PDIS items—item 11 (Bend over difficulty), 13 (Squat down difficulty), and 15 (Exercise difficulty).

Item-to-Item Correlations

Correlation between PDSS and PDIS item scores were sufficiently high ($r > 0.3$) to enable identification of scales in both instruments (eFigure 2, links.lww.com/CPJ/A445).

For the PDSS, 2 pairs of items—item 5 (Muscle weakness anywhere)/item 6 (Muscle weakness lower body) and item 10 (Muscle ache)/item 11 (Pain)—had strong correlations ($r = 0.90$ and $r = 0.91$, respectively), suggesting possible redundancy of items. Item 1 (Breathing)/item 2 (Breathing while lying down) also had high correlations ($r = 0.86$), as

did Item 3 (Tiredness)/Item 4 (Fatigue; $r = 0.89$). Many groups of items show consistently high Spearman rank correlations (>0.5).

For the PDIS, correlations suggest 2 scales, which match its intended design. The first included items 1, 2, and 3 (Anxiety, Worry, and Depression, respectively), and their Spearman rank correlations varied between $r = 0.78$ and $r = 0.93$. The second included the difficulty items (items 5, 7, 9, 11, 13, and 15 [Walk difficulty, Climb difficulty, Rise difficulty, Bend over difficulty, Squat down difficulty, and Exercise difficulty, respectively]). Correlations varied between $r = 0.46$ and $r = 0.80$.

Exploratory Factor Analysis

Results of exploratory factor analysis for the PDSS items indicated that a four-factor solution seemed to be optimal. The 4 factors addressed the underlying concepts of Shortness of Breath, Overall Fatigue, Pain, and Upper Extremity Weakness Score. Item 12 (Morning headache) did not load onto any of the 4 factors (factor loading <0.40 , Figure 1A). This item was, however, retained as a single-item scale, Morning Headache Score, given its stated relevance to patients during preliminary studies for the definition of the PDSS instrument.

Factor 1, labeled Overall Fatigue Score, accounted for most variability in the data. It was formed by 5 items: items 3 (Tiredness), 4 (Fatigue), 5 (Muscle weakness anywhere), 6 (Muscle weakness lower body), and 9 (Muscle weakness upper body). All standardized factor loadings were quite high (>0.8), except item 9, for which it was 0.54 (Figure 1A). The 2 items loading (>0.7) onto factor 2 relate to upper extremity weakness: items 7 (Muscle weakness arms) and 8 (Muscle weakness hand); this factor was labeled Upper Extremity Weakness Score. Two items related to pain (with loadings >0.8) were loaded on factor 3; this was labeled Pain Score. Finally, the 2 items relating to breathing, items 1 and 2 (Breathing and Breathing while lying down, respectively), with loadings >0.9 , were loaded on factor 4, which was labeled Shortness of Breath Score. Given the moderate-to-strong correlations ($r = 0.45$ to 0.62) observed between factor 1 (Overall Fatigue), factor 2 (Upper Extremity Weakness Score), and factor 3 (Pain Score), we hypothesized that a second-order dimension labeled Fatigue/Pain Score, underlying the initial factors 1 to 3 and indirectly items 3 to 11, is justifiable.

For the PDIS, analysis defined a two-factor solution. All standardized factor loadings were quite high (>0.7 , Figure 1B). Three items related to anxiety, worry, and depression were loaded on factor 1, which was labeled Mood Score. Items 5 (Walk difficulty), 7 (Climb difficulty), 9 (Rise difficulty), 11 (Bend over difficulty), and 13 (Squat down difficulty) were loaded on factor 2, which was labeled Difficulty Performing Activities Score. Item 15 (Exercise difficulty) was excluded from the analysis because few patients reported exercising, and therefore, the difficulty item was not completed.

Table 1 Pompe Disease Symptom Scale and Pompe Disease Impact Scale at Baseline—Item Score Distribution

	Item	Description (response option score)	Valid responses, N	Mean (SD)	Median (IQR)	Ceiling effect, ^a n (%)	Floor effect, ^b n (%)
PDSS	1	Breathing (0–10)	72	2.4 (2.0)	1.9 (1.0–3.7)	14 (19.4)	0 (0.0)
	2	Breathing while lying down (0–10)	72	2.6 (2.3)	2.0 (0.6–3.9)	15 (20.8)	0 (0.0)
	3	Tiredness (0–10)	72	4.0 (2.0)	3.8 (2.4–5.6)	0 (0.0)	0 (0.0)
	4	Fatigue (0–10)	72	3.9 (2.1)	3.7 (2.2–5.6)	0 (0.0)	0 (0.0)
	5	Muscle weakness anywhere (0–10)	72	4.6 (2.1)	4.6 (2.9–6.0)	2 (2.8)	1 (1.4)
	6	Muscle weakness lower body (0–10)	72	4.8 (2.1)	5.0 (3.1–6.0)	2 (2.8)	1 (1.4)
	7	Muscle weakness arms (0–10)	72	2.6 (2.0)	2.4 (1.4–3.2)	7 (9.7)	1 (1.4)
	8	Muscle weakness hand (0–10)	72	1.7 (2.1)	1.1 (0.0–2.7)	19 (26.4)	1 (1.4)
	9	Muscle weakness upper body (0–10)	72	4.0 (2.2)	4.0 (2.6–5.6)	3 (4.2)	1 (1.4)
	10	Muscle aches (0–10)	72	3.7 (2.4)	3.9 (1.6–5.3)	6 (8.3)	0 (0.0)
	11	Pain (0–10)	72	3.8 (2.3)	3.5 (2.0–5.6)	5 (6.9)	0 (0.0)
	12	Morning headache (0–10)	72	1.1 (1.5)	0.5 (0.0–1.6)	27 (37.5)	0 (0.0)
PDIS	1	Anxiety (0–10)	71	2.2 (1.9)	2.1 (0.4–3.5)	14 (19.7)	0 (0.0)
	2	Worry (0–10)	71	2.5 (2.1)	2.5 (0.4–3.5)	9 (12.7)	0 (0.0)
	3	Depression (0–10)	71	1.5 (1.7)	1.0 (0.0–2.4)	24 (33.8)	0 (0.0)
	5	Walk difficulty (0–4)	69	1.6 (0.9)	1.5 (1.0–2.0)	6 (8.7)	2 (2.9)
	7	Climb difficulty (0–4)	60	2.2 (1.0)	2.0 (1.4–2.9)	1 (1.7)	8 (13.3)
	9	Rise difficulty (0–4)	71	2.0 (1.0)	2.0 (1.0–2.7)	1 (1.4)	7 (9.9)
	11	Bend over difficulty (0–4)	63	2.1 (1.3)	1.9 (1.0–3.1)	2 (3.2)	12 (19.0)
	13	Squat down difficulty (0–4)	60	3.0 (1.1)	3.6 (2.0–4.0)	1 (1.7)	25 (41.7)
15	Exercise difficulty (0–4)	48	3.1 (1.2)	4.0 (2.0–4.0)	1 (2.1)	26 (54.2)	

Abbreviations: IQR = interquartile range; PDIS = Pompe Disease Impact Scale; PDSS = Pompe Disease Symptom Scale.

The PDSS is an 11-point scale from 'none' (0) to 'as bad as I can imagine' (10). Higher scores indicate greater severity of symptoms.

Items 1–3 of the PDIS are rated using an 11-point scale from 'no impact' (0) to 'as bad as I can imagine' (10). In the PDIS, for difficulty performing activities, there are 6 pairs of questions. The first of each pair asks whether the activity was performed in the past 24 h (items 4, 6, 8, 10, 12, and 14 [not shown here]), with patients answering 'no and not physically able' (0), 'no but physically able' (1), or 'yes' (2). Patients answering 'yes' are asked the follow-up question in the pair, categorizing the degree of difficulty in performing the activity in the past 24-h (items 7, 9, 11, 13, and 15). This follow-up question is on a 5-point scale: 'not at all difficult' (0), 'a little difficult' (1), 'somewhat difficult' (2), 'very difficult' (3), and 'extremely difficult' (4). Patients responding with '0' or '1' on the first question do not go on to answer the second question. Higher scores indicate greater impact on items.

^a Proportion of participants responding at the least severe end of the response scale.

^b Proportion of participants responding at the most severe end of the response scale.

In all cases, the items correlated strongly with their own scales and less with the other scales (eFigure 3, links.lww.com/CPJ/A445).

For both the PDSS and PDIS, scores were calculated as the average of underlying items. Therefore, items and scales shared the same range, 0 to 4 for Difficulty Performing Activities Score and 0 to 10 for all other scales. Higher scores indicate greater severity of symptoms or larger impact.

Reliability

All scales from the PDSS and PDIS showed excellent internal consistency with Cronbach alpha coefficient >0.90, ranging between 0.92 and 0.95 for the PDSS and 0.91 and 0.93 for the PDIS (Table 2).

Test-retest reliability was also demonstrated (Table 3). Both PDSS and PDIS domains and total scales had adequate-to-very high ICC values. All ICC values were >0.7 at screening and ranged from 0.60 to 0.85 between baseline and week 49, indicating acceptable test-retest reliability for the scale scores in both instruments.

Construct Validity

Convergent validity was demonstrated by moderate-to-high correlations of PDSS and PDIS scale scores with concepts hypothesized to be similar from the other PRO and clinical measures and especially those with the most similar concepts (Figures 2, A and B, respectively). Discriminant validity was supported by the observation of lower (low-to-moderate) correlations with the other PRO measures, which captured less similar concepts.

Figure 1 Exploratory Factor Analysis



(A) Pompe Disease Symptom Scale; (B) Pompe Disease Impact Scale.

Sensitivity to Change

Sensitivity to change for the PDSS was supported by logical differences in change from baseline to week 49 among PGIC change

groups, with significant differences observed for Total Symptom Score ($p = 0.031$), Fatigue/Pain Score ($p = 0.011$), and Overall Fatigue Score ($p = 0.014$; eTable 2, links.lww.com/CPJ/A445).

Table 2 Internal Consistency Reliability

	Description	Cronbach alpha (95% confidence interval)
PDSS	Total Symptom Score	0.92 (0.89–0.95)
	Shortness of Breath Score	0.93 (0.88–0.95)
	Fatigue/Pain Score	0.94 (0.92–0.96)
	Overall Fatigue Score	0.95 (0.93–0.97)
	Upper Extremity Weakness Score	0.92 (0.87–0.95)
	Pain Score	0.95 (0.92–0.97)
PDIS	Mood Score	0.93 (0.90–0.95)
	Difficulty Performing Activities Score	0.91 (0.86–0.95)

Abbreviations: PDIS = Pompe Disease Impact Scale; PDSS = Pompe Disease Symptom Scale.

The magnitude of these change scores were in the region of moderate-to-high effect sizes for the improved category. The results were inconclusive for detecting worsening owing to the small magnitude of change from baseline and lower sample size.

Similarly, sensitivity to change for the PDIS instrument was demonstrated by the logical differences in change from baseline to week 49 among PGIC change groups and small (for worsened/no change group) to moderate (improved group) effect sizes (eTable 2, links.lww.com/CPJ/A445).

Meaningful Change Thresholds

Candidate within-patient meaningful threshold values were generally identified from the anchor-based analysis as the lowest median value for a PGIC improvement category that exceeded the distribution-based approach estimates

and the 95% confidence interval (CI) for the stable (i.e., ‘no change’) anchor group. The correlations between changes from baseline to week 49 on the PDSS/PDIS scores and PGIC responses were used to evaluate the reliability of the meaningful change estimates. PGIC items, yielding correlations ≥ 0.30 with the PDSS/PDIS scores, were used as the primary anchors in estimating the thresholds. Estimates were rounded to a single decimal place.

PDSS

The Overall Fatigue, Fatigue/Pain, and Shortness of Breath Scores were correlated ≥ 0.30 with at least one PGIC item; the Pain, Upper Extremity Weakness, and Morning Headache Scores were not correlated with any of the PGIC items ≥ 0.30 ; and therefore, the confidence in the meaningful change estimates is reduced for these domains (Figure 3). Because the correlations between the PDSS scores and PGIC item 3 (ability to breathe) were generally stronger than other PGIC items, estimates based on this anchor were highlighted.

Table 4 includes the mean and median change score values for the PDSS scores across the PGIC response categories. Patients are categorized as having experienced “no change,” “improvement” (any response indicating improvement on the PGIC), or “large improvement” (moderate or large improvement selected on PGIC) based on their responses to each PGIC item.

The smallest median change value on the Shortness of Breath domain that exceeded the distribution-based estimates and the lower bound of the correlated PGIC items was -1.47 . For the ability to breathe PGIC anchor item, this value was -1.16 . For the Fatigue/Pain Score, the mean and median values exceeded the lower bound of the 95% CI for the stable group on 3 of the 4 PGIC items. For the ability to breathe PGIC item, the median value for the “large improvement” group (-1.30) exceeded the

Table 3 Test-Retest Reliability

	Scale	N	Test: day -14 to day -8 Retest: day -7 to day -1	N	Test: baseline (day -7 to day -1) Retest: week 49
PDSS	Total Symptom Score	41	0.88 (0.78, 0.93)	8	0.83 (0.33, 0.96)
	Shortness of Breath Score	41	0.88 (0.79, 0.93)	12	0.84 (0.54, 0.95)
	Fatigue/Pain Score	41	0.90 (0.82, 0.95)	8	0.84 (0.43, 0.97)
	Weakness in Hand Grip Score	41	0.89 (0.80, 0.94)	13	0.73 (0.34, 0.91)
	Morning Headache Score	41	0.77 (0.61, 0.87)	8	0.63 (-0.05, 0.91)
	Overall Fatigue Score	41	0.91 (0.83, 0.95)	8	0.85 (0.45, 0.97)
	Upper Extremity Weakness Score	41	0.90 (0.82, 0.94)	8	0.60 (-0.03, 0.90)
	Pain Score	41	0.87 (0.77, 0.93)	8	0.81 (0.33, 0.96)
PDIS	Mood Score	41	0.84 (0.73, 0.91)	13	0.67 (0.22, 0.89)
	Difficulty Performing Activities Score	38	0.96 (0.92, 0.98)	12	0.81 (0.48, 0.94)

Abbreviations: PDIS = Pompe Disease Impact Scale; PDSS = Pompe Disease Symptom Scale.

Table 4 Clinically Meaningful Thresholds (Anchor-Based and Distribution Method)—Large/Any Improvement Groups [Mean/Median] vs No Change Group [95% Confidence Intervals]

Total or domain	Groups	Mean/median for improvement groups vs [95% confidence intervals] no change				½ SD	1 SEM	
		PGIC 1 Ability with daily activities	PGIC 2 Disease-related symptoms	PGIC 3 Ability to breathe	PGIC 4 Mobility			
PDSS	Shortness of Breath Score	Large improvement	-1.80/-1.47 ^a	-1.12/-1.00	-1.86/-1.68 ^a	-1.70/-2.23	1.05	0.57
		Any improvement	-1.01/-0.95	-0.95/-0.88	-1.16/-0.94	-1.13/-0.90		
		No change	[-1.20, 0.34]	[-1.47, 0.34]	[-0.96, 0.57]	[-1.08, 0.66]		
	Fatigue/Pain Score	Large improvement	-1.30/-1.12 ^a	-1.16/-0.98 ^a	-1.56/-1.30 ^a	-1.46/-1.28 ^a	0.89	0.43
		Any improvement	-1.09/-0.94	-0.93/-0.68	-0.98/-0.71	-1.02/-0.83		
		No change	[-0.79, 0.36]	[-1.60, 0.36]	[-1.03, 0.17]	[-1.29, 0.17]		
Overall Fatigue Score	Large improvement	-1.42/-1.27 ^a	-1.37/-1.00 ^a	-1.85/-1.58 ^a	-2.03/-1.61 ^a	0.96	0.42	
	Any improvement	-1.24/-0.76	-1.08/-0.71	-1.18/-0.75	-1.21/-0.83			
	No change	[-0.87, 0.37]	[-1.59, 0.43]	[-0.98, 0.44]	[-1.29, 0.33]			
Pain Score	Large improvement	-1.51/-1.85	-1.30/-1.53	-1.39/-1.17	-1.03/-0.20	1.15	0.51	
	Any improvement	-1.20/-0.71	-0.97/-0.49	-0.96/-0.52	-1.03/-0.42			
	No change	[-0.71, 0.84]	[-1.77, 1.08]	[-1.38, -0.15]	[-1.53, 0.41]			
Upper Extremity Weakness Score	Large improvement	-0.81/-0.47	-0.49/-0.30	-0.99/-0.60	-0.47/-0.25	1.00	0.56	
	Any improvement	-0.64/-0.22	-0.50/-0.19	-0.51/-0.13	-0.53/-0.13			
	No change	[-1.36, 0.55]	[-2.61, 0.65]	[-1.68, 0.68]	[-1.61, 0.15]			
Morning Headache Score	Large improvement	-0.67/-0.28	-0.38/0.00	-0.85/-0.26	-0.59/-0.54	0.73	0.70	
	Any improvement	-0.47/0.00	-0.35/0.00	-0.35/0.00	-0.42/0.00			
	No change	[-1.36, 0.44]	[-2.44, 0.45]	[-1.53, 0.70]	[-1.40, 0.23]			
PDIS	Mood Score	Large improvement	-1.24/-0.85	-0.92/-0.42	-1.75/-1.61 ^a	-1.70/-2.01	0.88	0.47
		Any improvement	-0.91/-0.42	-0.70/-0.29	-0.86/-0.48	-0.92/-0.51		
		No change	[-0.90, 0.46]	[-1.67, 0.13]	[-1.46, 0.33]	[-1.08, 0.29]		
Difficulty Performing Activities Score	Large improvement	-0.71/-1.00 ^a	-0.69/0.91 ^a	-0.36/-0.25	-0.59/-0.95 ^a	0.48	0.28	
	Any improvement	-0.51/-0.57	-0.44/-0.39	-0.42/-0.31	-0.45/-0.40			
	No change	[-0.46, 0.13]	[-0.70, 0.23]	[-0.72, 0.11]	[-0.71, 0.00]			

Abbreviations: ½ SD = one-half standard deviation; PGIC = Patient Global Impression of Change; SEM = standard error of measurement.

Large improvement categories: 'a great deal better' and 'moderately better.' Improvement categories: 'a great deal better,' 'moderately better,' and "somewhat better."

^a Correlation between PDSS/PDIS score and PGIC item ≥0.30.

Figure 2 Construct Validity: Convergent and Divergent Validity

A

PRO	Description	TSS	SBS	FPS	OFS	PS	UEWS	MHS
EQ-5D-5L	Pain	(70) 0.451	(70) 0.057	(70) 0.525	(70) 0.403	(70) 0.717	(70) 0.338	(70) 0.130
	Mobility	(70) 0.474	(70) 0.144	(70) 0.502	(70) 0.469	(70) 0.401	(70) 0.431	(70) -0.031
SF-12	PCS	(70) -0.444	(70) -0.207	(70) -0.467	(70) -0.403	(70) -0.457	(70) -0.292	(70) -0.003
R-PAct	Total	(31) -0.463	(31) -0.033	(31) -0.559	(31) -0.460	(31) -0.490	(31) -0.420	(31) 0.113
Forced vital capacity		(53) -0.024	(53) -0.410	(53) 0.067	(53) 0.022	(53) 0.163	(53) -0.031	(53) 0.236
Six-minute walk test		(71) -0.184	(71) -0.028	(71) -0.215	(71) -0.196	(71) -0.150	(71) -0.222	(71) 0.156
Quick motor function test		(68) -0.181	(68) -0.087	(68) -0.197	(68) -0.251	(68) 0.031	(68) 0.165	(68) 0.266
PDIS	MS	(71) 0.602	(71) 0.325	(71) 0.591	(71) 0.631	(71) 0.320	(71) 0.486	(71) 0.255
	DPAS	(70) 0.492	(70) 0.192	(70) 0.518	(70) 0.535	(70) 0.319	(70) 0.353	(70) -0.025

B

PRO	Description	MS	DPAS
EQ-5D-5L	Pain	(70) 0.091	(69) 0.306
	Mobility	(70) 0.172	(69) 0.538
	Anxiety and depression	(70) 0.384	(69) 0.298
SF-12	MCS	(70) -0.407	(69) -0.220
	PCS	(70) -0.020	(69) -0.486
R-PAct	Total	(31) -0.239	(31) -0.669
Forced vital capacity		(53) -0.161	(53) -0.010
Six-minute walk		(71) -0.081	(70) -0.454
Quick motor function		(68) -0.167	(67) -0.606
PDSS	TSS	(71) 0.602	(70) 0.492
	SBS	(71) 0.325	(70) 0.192
	FPS	(71) 0.591	(70) 0.518
	OFS	(71) 0.631	(70) 0.535
	PS	(71) 0.320	(70) 0.319
	UEWS	(71) 0.486	(70) 0.353
	MHS	(71) 0.255	(70) -0.025

(A) Pompe Disease Symptom Scale; (B) Pompe Disease Impact Scale. Data are (n) Spearman rank correlations. Dark green = strong correlations ($r > 0.5$); light green = moderate correlations ($r = 0.3$ to 0.5); yellow = weak correlations ($r < 0.3$). DPAS = Difficulty Performing Activities Score; EQ-5D-5 = EuroQol-5 dimensions 5 levels; FPS = Fatigue/Pain Score; MCS = Mental Component Summary; MHS = Morning Headache Score; MS = Mood Score; OFS = Overall Fatigue Score; PCS = Physical Component Summary; PS = Pain Score; R-PAct = Rasch-built Pompe-specific activity; SBS = Shortness of Breath; SF-12 = Short Form 12 Items; TSS = Total Symptom Score; UEWS = Upper Extremity Weakness Score.

lower bound of the 95% CI. A similar trend was observed for the Overall Fatigue Score. For each of these scores, a meaningful change threshold of -1.5 was selected because this value generally exceeded the lower bound of the 95% CIs, exceeded the distribution-based estimates, and was consistent with the changes observed for the “large

improvement” group across the PGIC items. The other PDSS scores did not correlate with the PGIC anchors. Therefore, a meaningful change estimate was selected that exceeded the distribution-based methods and was consistent with the thresholds for the other scores. A threshold of -1.5 was selected for these scores as well.

Figure 3 Correlation Between Change Score in PDSS and PDIS Scales and PGIC Anchor Items, Baseline to Week 49

Total/domain scores	PGIC Item 1 Daily activities	PGIC Item 2 Disease related	PGIC Item 3 Breath	PGIC Item 4 Mobility	
PDSS	Shortness of breath score	-0.317	-0.244	-0.471	-0.260
	Fatigue/pain score	-0.365	-0.374	-0.486	-0.364
	Overall fatigue score	-0.368	-0.416	-0.541	-0.455
	Upper extremity weakness score	-0.291	-0.171	-0.142	-0.105
	Pain score	-0.242	-0.169	-0.201	-0.082
	Morning headache score	0.057	0.228	-0.188	0.110
PDIS	Mood score	-0.182	-0.098	-0.441	-0.246
	Difficulty performing activities score	-0.483*	-0.486*	-0.170*	-0.339*

Dark green: strong correlation, light green: moderate correlations, light yellow: weak correlations. N = 50 unless indicated otherwise. *N = 46. PDIS = Pompe Disease Impact Scale; PDSS = Pompe Disease Symptom Scale; PGIC = Patient Global Impression of Change.

PDIS

The Mood Score was correlated by ≥ 0.30 with the ability to breathe PGIC item while the Difficulty Performing Activities Score was correlated by ≥ 0.30 with the daily activities, disease-related symptoms, and mobility PGIC items (Figure 3). The PGIC items with stronger correlations with the PDIS scores were emphasized in estimating the meaningful change thresholds.

For the Mood Score, the median value of -1.61 from the “large improvement” group exceeded the lower bound of the 95% CI for the stable group on the ability to breathe PGIC item. This estimate also exceeded the distribution-based estimates and the lower bound of the 95% CI for the stable groups for 2 of the other 3 PGIC items. To be consistent with the thresholds for the PDSS, a threshold of -1.5 was selected for the Mood Score.

For the Difficulty Performing Activities Score, the scale range is smaller (0–4) than the other scores (0–10). Therefore, the meaningful change estimate is also smaller than for the other scores. The lower bound of the 95% CI for the stable group was approximately -0.70 for 3 of the 4 PGIC items. The smallest value that exceeded this was the median value of the “large improvement” group across the PGIC items with stronger correlations with the Difficulty Performing Activities Score; this estimate was approximately -1.0 . This value also exceeded the distribution-based values and was selected as the meaningful change estimate for this score.

Discussion

Patients with LOPD experience a variety of symptoms and functional limitations that can substantially decrease quality of life. Reliable and valid measures of how patients with LOPD experience their condition are needed for use in clinical research. The items of the PDSS and PDIS were generated through rigorous qualitative research³ and are designed to quantify the patient’s perspective on their symptoms and functioning. Data from the COMET clinical trial support the psychometric properties of these novel PRO measures.

The PDSS yields scores assessing key symptom domains of LOPD, including Shortness of Breath, Overall Fatigue, Pain, Upper Extremity Weakness, and Morning Headache. The Overall Fatigue domain includes aspects of both tiredness and muscle weakness in different areas of the body. Although these could be considered as separate concepts, factor analysis indicated that they were highly correlated and measured the same underlying experience. Therefore, they were combined into a single domain score. Because degeneration of muscle strength is a key clinical feature of LOPD, this specific domain score may be particularly relevant in future research with LOPD. Additional PDSS items evaluating specific dimensions of upper extremity weakness—weakness in the

arms and hands—formed a separate domain, rather than loading with the more general muscle weakness item asking about “upper body” weakness. This separate domain may supplement the Overall Fatigue domain in future research because it measures a specific symptom that may be important in fully understanding the patient’s experience of LOPD symptoms.

The Overall Fatigue and Pain domains were strongly correlated. Although they measure separate concepts, they also seemed to capture a common underlying symptom cluster. Therefore, a second-order factor was retained that encompassed these symptoms. This composite domain may provide an efficient method of indexing these different experiences using a single score in future clinical research. Morning headache was not closely associated with any of the other items, and a notable floor effect was observed for this item. Because the item captures a potentially important aspect of the patient’s experience, it was retained as a stand-alone score that is not combined with any of the other items. However, it may be considered for deletion in future versions of the PDSS.

The PDIS includes 2 domain scores—a Mood Score and Difficulty Performing Activities Score. The Mood Score includes ratings on the 3 items measuring the severity of negative moods—anxiety, depression, and worry. The Difficulty Performing Activities Score addresses basic instrumental activities of daily living that can be affected by muscle weakness, fatigue, and shortness of breath. Difficulty with exercising was omitted from the Difficulty Performing Activities scale because it seemed to measure a concept that is much less commonly attempted or achievable by the target population. Nevertheless, examining it as a separate index may be useful in LOPD samples without profound physical limitations. The dichotomous items addressing whether the patient completed an activity are also not included in the Difficulty Performing Activities scale. Nevertheless, these items may provide useful information on their own, in that moving from not being able to complete an activity to being able to complete it may be an important treatment benefit. Alternatively, losing function and changing over time from being able to complete an activity to not being able to complete the activity may be an important marker of disease progression.

The reliability and validity of the PDSS and PDIS were supported. The correlations with the criterion measures indicated that the symptom and impact ratings were strongly associated with other PRO measures of symptoms, functioning, and health status. In addition, there were specific relationships with pulmonary, exercise capacity, and motor function measures that strongly supported the convergent and discriminant validity of the PDSS and PDIS: The Shortness of Breath Score of the PDSS was associated with the FVC pulmonary function measure, and the Difficulty Performing Activities Score was associated with the 6MWT

and QMFT. These specific correlations differentiated the Shortness of Breath Score and Difficulty Performing Activities Score from the other domains, as expected. Similarly, although the Pain item of the EQ-5D was correlated with several PDSS domains, it was most strongly related to the PDSS Pain Score, as expected. The pattern of change scores from baseline to week 49 were generally ordered correctly, where patients who reported improvement on a global impression item also reported greater reductions in symptoms and increased ability to perform activities. Only the Mood Score was disordered. However, the anchor items did not assess mood—they included symptoms and physical limitations—so this result is not surprising. In general, this pattern of convergent and divergent results strongly supports the construct validity of the PDSS and PDIS domain scores. The structural validity of the instruments would ideally be confirmed eventually on a different, larger sample.

Establishing a meaningful change threshold is an important step in developing a new measure for use in clinical research or practice. This allows each patient to be categorized as having experienced meaningful improvement or worsening in their symptoms or function. The analyses reported here followed the best practice guidelines of methodological experts and regulatory agencies. The estimates were generally consistent across the PDSS and PDIS domains. A decrease of 1 to 1.5 points or greater indicates a meaningful improvement on the PDSS and PDIS domains. These thresholds can be used in clinical research studies to identify responders to treatment. The relatively low correlations between the anchors and target measures increase the uncertainty around the meaningful change point estimates. When applying these thresholds, it may be useful to consider ranges of thresholds, and not just the single point estimates, to ensure that a robust change has occurred for individual patients.

The PDSS and PDIS were developed using regulatory and expert best practice guidelines.²¹ This iterative process for PRO development included qualitative interviews, disease-specific questionnaire development, and quantitative psychometric analyses. The PDSS and PDIS are reliable and valid measures that can be used to evaluate important LOPD-specific symptoms and impacts on patients in observational research and clinical trials and to monitor disease progression in clinical practice.

Acknowledgment

The authors acknowledge administrative assistance from Jane M. Gilbert, BSc, CMPP, and Kim Coleman Healy, PhD, CMPP, of Envision Scientific Solutions, contracted by Sanofi for publication support services. The authors thank Leah Granby, Naveen Deenadayalu, Kristen Gyorda, Susan Macera, Megan Cirlincione, and Olivier Huynh-Ba for their contributions. The authors exerted sole scientific control and received no honoraria for participation. Coauthor Giulio Flore, MSc, died on November 4, 2022.

Study Funding

This study was supported by Sanofi.

Disclosure

M.M. Dimachkie has received consulting fees from Abata/Third Rock, Abcuro, Amazentis, ArgenX, Astellas, Catalyst, Cello, Covance/Labcorp, CSL-Behring, EcoR1, Janssen, Kezar, MDA, Medlink, Momenta, NuFactor, Octapharma, Priovant, RaPharma/UCB, Roivant Sciences Inc, Sanofi, Scholar Rock, Shire Takeda, Spark Therapeutics, TACT, UCB Biopharma, and UpToDate. M.M. Dimachkie has undertaken contracted research for or received unrestricted educational grants from Alexion, Alnylam Pharmaceuticals, Amicus, Biomarin, Bristol-Myers Squibb, Catalyst, Corbus, CSL-Behring, FDA/OOPD, GlaxoSmithKline, Genentech, Grifols, Kezar, Mitsubishi Tanabe Pharma, MDA, NIH, Novartis, Octapharma, Orphazyme, Ra Pharma/UCB, Sanofi, Sarepta Therapeutics, Shire Takeda, Spark Therapeutics, The Myositis Association, TMA, UCB Biopharma/RaPharma, and Viomed/Healixmith. P.S. Kishnani has received research/grant support from Sanofi and Amicus Therapeutics and has received consulting fees and honoraria from Sanofi, Amicus Therapeutics, Maze Therapeutics, JCR Pharmaceutical, and Asklepios Biopharmaceutical, Inc. (AskBio). P.S. Kishnani is a member of the Pompe and Gaucher Disease Registry Advisory Boards for Sanofi, Amicus Therapeutics, and Baebies. P.S. Kishnani has equity in Asklepios Biopharmaceutical, Inc. (AskBio), and Maze Therapeutics. C. Ivanescu is an employee of IQVIA. G. Flore was an employee of IQVIA at the time of study participation. C. Gwaltney has received consulting fees from Sanofi. N.A.M.E. van der Beek has received consulting fees and travel reimbursement from Sanofi and Amicus Therapeutics. A. Hamed is an employee of Sanofi. K. An Haack and her spouse are employees of Sanofi. L. Pollissard is an employee of Sanofi. E. Baranowski was an employee of Sanofi at the time of study participation. S.E. Sparks is an employee of Sanofi. P. DasMahapatra is an employee of Sanofi. Full disclosure form information provided by the authors is available with the full text of this article at [Neurology.org/cp](https://www.neurology.org/cp).

Publication History

Received by *Neurology: Clinical Practice* January 19, 2023. Accepted in final form June 9, 2023. Submitted and externally peer reviewed. The handling editor was Associate Editor Belinda A. Savage-Edwards, MD, FAAN.

Appendix Authors

Name	Location	Contribution
Mazen M. Dimachkie, MD,	University of Kansas Medical Center, Kansas City, KS, USA	Data acquisition and interpretation, and revising of the manuscript
Priya S. Kishnani, MD	Duke University Medical Center, Durham, NC, USA	Data acquisition and interpretation, drafting and revising of the manuscript

Continued

Appendix (continued)

Name	Location	Contribution
Cristina Ivanescu, PhD	IQVIA, Patient Centered Solution, Amsterdam, Netherlands	Psychometric analysis plan development, psychometric analysis and data interpretation, drafting and revising of the manuscript, statistical analysis
Giulio Flore, MSc	IQVIA, Patient Centered Solution, Amsterdam, Netherlands	Psychometric analysis plan development, psychometric analysis and data interpretation, drafting and revising of the manuscript
Chad Gwaltney, PhD	Gwaltney Consulting, Westerly, RI, USA	Study concept and design, statistical analysis, data interpretation, drafting and revising of the manuscript
Nadine A.M.E. van der Beek, MD, PhD	Center for Lysosomal and Metabolic Diseases, Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands	Data acquisition and interpretation, and revising of the manuscript
Alaa Hamed, MD, MPH, MBA	Sanofi, Cambridge, MA, USA	Study concept and design, data acquisition and interpretation, drafting and revising of the manuscript
Kristina An Haack, MD	Sanofi, Chilly-Mazarin, France	Study concept and design, data interpretation, drafting and revising of the manuscript
Laurence Pollissard, MSc	Sanofi, Chilly-Mazarin, France	Data interpretation, drafting and revising of the manuscript
Eileen Baranowski, PharmD, MS	Sanofi, Cambridge, MA, USA	Psychometric analysis plan development, statistical analysis, data interpretation, drafting and revising of the manuscript
Susan E. Sparks, MD, PhD	Sanofi, Cambridge, MA, USA	Data interpretation, drafting and revising of the manuscript
Pronabesh DasMahapatra, MD, MPH	Sanofi, Cambridge, MA, USA	Study concept and design, data acquisition and interpretation, drafting and revising of the manuscript

References

- Reuser A, Hirschhorn R, Kroos M. Part 16: Lysosomal disorders: Pompe disease: glycogen storage disease Type II, acid α -glucosidase (acid Maltase) deficiency. In: Valle D, Antonarakis S, Ballabio A, Beudet A, Mitchell G, eds. *OMMBID – The Online Metabolic and Molecular Bases of Inherited Diseases*. McGraw-Hill Education, 2019.
- van der Ploeg AT, Reuser AJ. Pompe's disease. *Lancet*. 2008;372(9646):1342-1353. doi:10.1016/s0140-6736(08)61555-x
- Hamed A, An Haack K, Gwaltney C, et al. Qualitative interviews to improve patient-reported outcome measures in late-onset Pompe disease: the patient perspective. *Orphanet J Rare Dis*. 2021;16(1):428. doi:10.1186/s13023-021-02067-x
- Diaz-Manera J, Kishnani PS, Kushlaf H, et al. Safety and efficacy of avalglucosidase alfa versus alglucosidase alfa in patients with late-onset Pompe disease (COMET): a phase 3, randomised, multicentre trial. *Lancet Neurol*. 2021;20(12):1012-1026. doi:10.1016/S1474-4422(21)00241-6
- McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293-307. doi:10.1007/BF01593882
- Meyers L, Gamst G, Guarino A. *Applied Multivariate Research: Design and Interpretation*. Sage, 2006.
- Fayers PM, Machin D. *Quality of Life. The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*. 2nd ed. Wiley, 2007.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
- Perrot S, Lanteri-Minet M. Patients' Global Impression of Change in the management of peripheral neuropathic pain: clinical relevance and correlations in daily practice. *Eur J Pain* 2019;23(6):1117-1128. doi:10.1002/ejp.1378
- Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233. doi:10.1097/00005650-199603000-00003
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x
- van der Beek NA, Hagemans ML, van der Ploeg AT, van Doorn PA, Merckies IS. The Rasch-built Pompe-specific activity (R-PAct) scale. *Neuromuscul Disord*. 2013;23(3):256-264. doi:10.1016/j.nmd.2012.10.024
- Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-338. doi:10.1183/09031936.05.00034805
- Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40(6):1324-1343. doi:10.1183/09031936.00080312
- Gibbons WJ, Fruchter N, Sloan S, Levy RD. Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. *J Cardiopulm Rehabil*. 2001;21(2):87-93. doi:10.1097/00008483-200103000-00005
- ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166(1):111-117. doi:10.1164/ajrccm.166.1.at1102
- Geiger R, Strasak A, Treml B, et al. Six-minute walk test in children and adolescents. *J Pediatr*. 2007;150(4):395-399. doi:10.1016/j.jpeds.2006.12.052
- van Capelle CI, van der Beek NA, de Vries JM, et al. The quick motor function test: a new tool to rate clinical severity and motor function in Pompe patients. *J Inherit Metab Dis*. 2012;35(2):317-323. doi:10.1007/s10545-011-9388-3
- Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care*. 2000;38(9 suppl 1):II84-II90.
- Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2(1):63-67. doi:10.1081/copd-200050663
- US Food and Drug Administration. Patient-focused drug development guidance public workshop. *Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments. Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 3. Workshop October 15-16, 2018* [online]. Accessed November 10, 2022. [fda.gov/downloads/Drugs/News-Events/UCM620708.pdf](https://www.fda.gov/downloads/Drugs/News-Events/UCM620708.pdf).

How to cite this article: Dimachkie MM, Kishnani PS, Ivanescu C, et al. Measurement properties of 2 novel PROs, the pompe disease symptom scale and pompe disease impact scale, in the COMET study. *Neurol Clin Pract*. 2023;13:e200181. doi: 10.1212/CPJ.000000000200181.