



Fabry Disease Patient-Reported Outcome (FD-PRO) demonstrates robust measurement properties for assessing symptom severity in Fabry disease

Alaa Hamed^a, Pronabesh DasMahapatra^a, Nicole Lyn^{a,*}, Chad Gwaltney^b, Charlie Iaconangelo^c, Daniel Serrano^c, Vijay Modur^d, Juan Politei^e

^a Sanofi Genzyme, Cambridge, MA, USA

^b Gwaltney Consulting, Westerly, RI, USA

^c Pharmerit International, Bethesda, MD, USA

^d Eloxx Pharmaceuticals, Watertown, MA, USA

^e Fundacion Para el Estudio de Las Enfermedades Neurometabolicas, Buenos Aires, Argentina

ARTICLE INFO

Keywords:

Fabry disease
Patient-reported outcome
Psychometrics
Validation
Lysosomal storage disorder
FD-PRO

ABSTRACT

Background: Fabry disease (FD) is a rare, genetic disease, that if untreated, progresses to irreversible and life-threatening renal, cardiac, and cerebrovascular events. FD symptoms impact daily functioning and quality of life, but no disease-specific measure of these symptoms has been psychometrically tested.

Methods: The Fabry Disease Patient-Reported Outcome (FD-PRO) consists of 19 items that measure neuropathic symptoms (pain, tingling, numbness and burning in upper/lower extremities), headache, abdominal pain, heat intolerance, swelling, tinnitus, fatigue, hearing/vision impairment, hypohidrosis (diminished sweating) and difficulty engaging in regular physical activities in the past 24 h. Measurement properties of the instrument were evaluated among 139 adult (≥ 18 years) FD diagnosed patients (enzyme deficiency in males; *GLA* genotyping in females) including enzyme replacement (ERT) treated or treatment-naïve patients, classic or late-onset phenotypes from ten countries and eighteen sites. Patients completed the FD-PRO daily on a handheld electronic diary for 4 weeks; demographic, other patient and clinician reported outcomes were also collected.

Results: The mean age of patients was 43 years; with even sex distribution (female: 53%) and majority was ERT treated (72%). Patient compliance was high; $\geq 87\%$ completed at least 4 FD-PRO entries each week (mean completion time: < 3 min in week one). Empirical evaluation of item properties via inter-item correlations, exploratory factor analysis and item-response theory models suggested that a total symptom score (TSS) could be calculated. Due to redundancy among items, a “neuropathy parcel” and an “audiovisual parcel” were created in generating the TSS (items within a parcel averaged and treated as a single item). Two items were excluded from TSS: sweating (did not correlate with other items) and difficulty engaging in regular physical activities (measure of impact, not symptoms). Internal consistency (Cronbach’s alpha) of the TSS was ≥ 0.89 across weeks; test-retest reliability (intraclass correlation coefficient) was ≥ 0.91 . The TSS was correlated with conceptually similar clinical and patient reported assessments as expected ($r > |0.4|$) and discriminated moderate/severe from least severe FD groups in known-groups validity analyses.

Conclusions: The FD-PRO instrument is a novel disease-specific instrument that assesses classic and non-classic symptoms, with strong psychometric properties and appropriate for use in clinical studies.

1. Background

Fabry disease (FD) is a rare, X-linked, genetic disorder caused by mutations in the galactosidase alpha (*GLA*) gene encoding for the

lysosomal enzyme, alpha-galactosidase A (α GAL).^{1,2} The α GAL deficiency caused by *GLA* mutations leads to progressive accumulation of globotriaosylceramide (GL3) in lysosomes of a variety of cell types.^{1,2} Clinically, there are two major subtypes: the early-onset, severe classic

* Corresponding author at: Sanofi Genzyme, 50 Binney Street, Cambridge, MA 02142, USA.

E-mail addresses: Alaa.Hamed@sanofi.com (A. Hamed), Pronabesh.DasMahapatra@sanofi.com (P. DasMahapatra), Nicole.Lyn@sanofi.com (N. Lyn), CGwaltney@gwaltneyconsulting.com (C. Gwaltney), CharlieIaconangelo@openhealthgroup.com (C. Iaconangelo), DanielSerrano@openhealthgroup.com (D. Serrano), Vmodur@eloxpharma.com (V. Modur), JPolitei@hotmail.com (J. Politei).

<https://doi.org/10.1016/j.ymgmr.2021.100824>

Received 7 October 2021; Accepted 14 November 2021

Available online 19 November 2021

2214-4269/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

phenotype or the late-onset phenotype. Systemic accumulation of GL-3 in capillary endothelial cells, cardiomyocytes and vascular smooth muscle cells podocytes plays a major role in the severe renal, cardiac, and cerebrovascular clinical manifestations, and decreased life expectancy in patients with FD.^{1–3} There is marked heterogeneity in clinical presentation and disease progression, with a variety of clinical manifestations that range from a multisystemic, severe form to milder phenotypes.⁴ Environmental, epigenetic, or modifier genes have been proposed as potential explanations for disease heterogeneity on the basis of their ability to modulate the clinical phenotype of individuals.⁴ Early in the course of the disease, symptoms of FD include neuropathic pain in the extremities, impaired sweating, heat intolerance, skin lesions, and gastrointestinal discomfort.² Such symptoms negatively impact quality of life of patients with FD and are a source of significant morbidity.²

To date, current treatment options include the reduction of accumulated glycosphingolipids with enzyme replacement therapy (agalsidase beta, agalsidase alfa) and pharmacologic chaperone (migalastat) approved for a subset of Fabry patients with amenable mutations. Although the ultimate goals of treatment for FD are preventing organ failure and death, reducing symptom burden and improving patients' quality of life is also an important clinical objective.¹ Symptom experience is best measured by asking patients to evaluate their own experiences using a disease-specific patient-reported outcome (PRO) measure.⁵ Such a measure provides insights into the clinical course of the disease, as well as used in clinical studies as an indicator of treatment efficacy and clinical benefit.

The Fabry Disease Patient-Reported Outcome (FD-PRO) was designed to capture symptom severity in patients with FD. The FD-PRO was developed based on a 2-step process: (1) **qualitative phase** that included a review of the published literature, and interviews with patients and clinical experts as the basis for initial item development and refinement; (2) **quantitative phase** to test the measurement properties of the items of the FD-PRO in an observational study (PROOF: An Observational Study to Assess the Psychometric Properties of a Patient-Reported Outcome [PRO] Instrument in Patients with Fabry Disease). Results from the qualitative phase of the instrument development are presented elsewhere.⁶ The current manuscript presents the measurement properties of the FD-PRO from analysis of the PROOF study. To our knowledge, this is the first disease-specific instrument in FD that has been empirically tested for validity, reliability and appropriateness of the instrument for clinical use.

2. Methods

2.1. Study design

Evaluation of the FD-PRO was based on a multicenter, international, prospective, longitudinal study (PROOF). This study was conducted in accordance with the guidelines for Good Epidemiology Practice^{7,8}. Each participating country was responsible for ensuring that all necessary regulatory submissions (IRB/IEC) were performed in accordance with local regulations, including local data protection regulations.

2.2. Study population

Eligible participants were adults aged ≥ 18 years with a confirmed diagnosis of FD (enzyme deficiency in males; *GLA* genotyping in females). Both classic or late-onset phenotypes were eligible, and patients were either treated by enzyme replacement therapy (ERT) or naïve to treatment at the time of enrollment. ERT-treated patients had to be stable at the recommended dose for the last 6 months. Patients with a history of organic disease other than FD including cardiovascular, hepatic, pulmonary, neurological, or renal disease were excluded.

2.3. Procedures

The FD-PRO instrument was administered daily via an electronic diary for 30 consecutive days (approximately 4 weeks). Patient Global Impression of Change (PGIC) and Patient Global Impression-Static (PGIS) was completed weekly via an electronic diary. Other patient-reported outcome (PROs) instruments and clinician-reported outcomes (ClinROs) were collected via paper in the clinic on Day 1 and Day 30 (week 4), while Estimated Meaningful Change in Symptoms Questionnaire and Stool Frequency and Consistency Questionnaire were collected via paper in the clinic only on Day 30 (week 4) (Table 1). The electronic diary was presented in patients' local language and questionnaires were translated and linguistically validated.

Age, sex, race, ethnicity, site location, medical history and medication list, and Fabry-specific treatment (ERT) were collected at screening. Phenotype was determined by collecting biomarker data at screening. For males, enzyme deficiency was determined via basophils/leukocyte or plasma a galactosidase A (α GAL) activity assay. For females, *GLA* genotype mutation was confirmed via genotyping. Characterization of classic or late-onset phenotype was based on characteristic *GLA* mutation as defined by the International Fabry Disease Phenotype-Genotype Database⁹ (both sex) or residual enzyme activity ($< 1\%$ of normal for classic males).

2.4. Measures

2.4.1. Patient-reported outcomes

FD-PRO is a 19-item PRO instrument to assess patient-reported symptoms associated with FD. Seventeen items assess presence and severity of each symptom using a numerical rating scale (NRS) ranging from 0 (none) to 10 (as bad as you can imagine). One item (yes/no/not sure) assesses whether the patient was in a situation that would have led to sweating. The final item assesses difficulty engaging in regular physical activities using a NRS ranging from 0 (no difficulty) to 10 (difficulty as bad as you can imagine). This latter item (physical activity) was not included in the scoring evaluation of the FD-PRO.

SF-36v2 is a 36-item generic PRO questionnaire used to assess patient-reported health related quality of life (HRQoL) outcomes.¹⁰ The items yield scores for 8 health domains—Physical Functioning (PF), Role Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role Emotional (RE), and Mental Health (MH) (higher scores reflect a better health state).

BDI-II consists of 21 self-reported items assessing the severity of depression symptoms, as described in American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders Fourth Edition using 5-point Likert-type response scales.

IBS-QoL is a 34-item PRO measure that assesses the impact of IBS on HRQoL using 5-point Likert-type response scales.¹¹ The individual responses to the 34 items are summed and averaged for a total score, and then transformed to a 0 to 100 scale for ease of interpretation, with higher scores indicating better IBS-specific HRQoL. There are also eight subscale scores for the IBS-QoL: Dysphoria, Interference with Activity, Body Image, Health Worry, Food Avoidance, Social Reaction, Sexual, Relationships.

DS3 patient item is a subsection of the DS3 instrument that consists of a single PRO item assessing overall patient well-being during the last month. It ranges from 0 (best overall well-being) to 4 (worst overall well-being).

PGIS is a single-item scale in which patients indicate a point-in-time, overall assessment of their FD symptoms (i.e., no FD symptoms, mild FD symptoms, moderate FD symptoms, severe FD symptoms, and very severe FD symptoms).

PGIC is a 20-item scale in which patients indicate the amount of improvement or worsening they have experienced, since the beginning of the study, on symptoms that approximately correspond to the symptoms evaluated in the FD-PRO.

Table 1
Schedule of assessments from enrollment to Day 30 ± 5 days (Week 4).

	Screening Period	Day 1	Observational Period			Study End (30 days ± 5 days)
Study Week	Day-14 to Day 1		1	2	3	4
Visit at clinical site	X					X
Informed consent	X					
Review Inclusion/Exclusion criteria including Symptom Screening Questionnaire (electronic diary)	X					
Demographics	X					
Medical History and Medications list ^a	X					
PRO Criterion Measures (to be completed by the patient)						
FD PRO Instrument including the BSFS (electronic diary) ^b			Daily assessment			
			←----->			
PGIS (electronic diary)		X	X	X	X	X
PGIC (electronic diary)			X	X	X	X
SF-36v2 (paper)		X				X
IBS-QOL (paper)		X				X
BDI-II (paper)		X				X
DS3 Fabry Disease Severity Scoring System (paper)		X				X
Estimated Meaningful Change in Symptoms Questionnaire (paper)						X
Stool Frequency and Consistency Questionnaire (paper)						X
ClinRO Criterion Measure (to be completed by the investigator)						X
DS3 Fabry Disease Severity Scoring System (paper) ^c		X				X
Physician Global Assessment (paper)		X				X

FD: Fabry Disease; PRO: Patient-Reported Outcome, FD-PRO: Fabry Disease Patient-Reported Outcome; BSFS: Bristol Stool Form Scale; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression-Static; SF-36v2: Short Form-36 Health Survey version 2; IBS-QOL: Irritable Bowel Syndrome Quality of Life; BDI-II: Beck Depression Inventory Second Edition; DS3: Fabry Disease Severity Scoring System; ClinRO: Clinician-reported outcome.

^a . Medications list is defined as medications taken by the patient at study entry collected by drug classes (according to WHO Anatomical Therapeutic Chemical (ATC) drug classification system).

^b . FD-PRO completed daily at bedtime (the last one at the night before the last visit).

^c . The Investigator completed all “clinical” domains of the D—S3 on Day 1 but only the Peripheral Nervous System (PNS) domain on Day 30.

Other PROs: As an additional measure of meaningful change, patients completed a survey that asked them to evaluate the hypothetical important and meaningful change in the FD-PRO score. Two item prompts asked the patient to specify meaningful improvement and deterioration relative to a rating of 6. Another two item prompts asked the patient to specify meaningful improvement and deterioration relative to a rating of 3.

Patients completed the **Stool Frequency and Consistency Questionnaire**, a single item question completed at the end of the study. The patient evaluated which was more bothersome, stool frequency or stool consistency.

Additionally, the **Bristol Stool Form Scale (BSFS)**, which asked patients to classify their bowel movements, was administered as part of the daily diary assessment.¹²

The **Symptom Screening Questionnaire** consisted of 11 items were administered at screening to evaluate the severity of burning, tingling, numbness of the extremities, diarrhea, abdominal pain, and tiredness during the past week. The scale, like the FD-PRO scale, ranged from 0 to 10, with 0 indicating no symptoms and 10 indicating symptoms as bad as can be imagined. The questionnaire was used to stratify patients into cohorts. Absent was defined as having a maximum score of 0 across all 11 items. Mild was defined as having a maximum score of 1-3 across all 11 items. Moderate was defined as having a maximum score of 4-7 across all 11 items. Severe was defined as having a maximum score of 8-11 across all 11 items.

2.4.2. Clinician-reported outcomes

Physician Global Assessment (PGA) is a single-item question regarding the physician’s global assessment of the patient’s signs and symptoms of their FD rated on a 6-point scale (e.g., 1 = no sign/symptoms, 2 = very mild signs/symptoms, 3 = mild signs/symptoms, 4 = moderate signs/symptoms, 5 = severe signs/symptoms, and 6 = very severe signs/symptoms).

DS3 is Fabry Disease Severity Scoring System that includes

Peripheral Nervous System (PNS) Domain (sweating, gastrointestinal, and pain subdomains), Renal Domain (estimated glomerular filtration rate [eGFR], proteinuria, and eGFR slope), Cardiac Domain (left ventricular hypertrophy [LVH], arrhythmia, and New York Heart Association Class of Heart Failure [NYHA] assessments), and Central Nervous System Domain (white matter lesions and transient ischemic attack [TIA]/stroke).

2.5. Statistical analysis

Sample Characterization: Sample demographics and clinical features were characterized by descriptive statistics. In the case of categorical variables, the sample size and percentage were reported, while the mean and standard deviation (SD) were reported for continuous variables.

Data Transformation and Missing Data Handling: The FD-PRO items were collected daily via electronic diary. The nature of the device ensured that patients who responded to the diary completed all items. That is, for a given diary, subjects either responded to all items or did not respond to any items. For purposes of analyzing the baseline data to empirically determine the domain structure, missing values were discarded because they contained no responses. Likewise, when scoring, there was no need to determine how to score cases with less than 100% missing data. Unless otherwise noted, daily data were averaged over seven days to create weekly scores for use in analyses of reliability, and validity. In order for a patient’s weekly score to be included in these analyses, at least four out of seven diaries must have been completed by the subject that week.

Compliance: Compliance rates were calculated to evaluate the willingness of patients to complete the daily diary on a weekly basis. Compliance for a given week was defined as the patient having completed a minimum of four daily diaries out of the seven possible. The rate of compliance was computed for weeks one through four. To evaluate patient burden, the average time required for patients to complete

the FD-PRO was also computed for each week.

FD-PRO Item Distributions: The weekly average of each item was computed for each patient and subsequently averaged across all patients to obtain the distribution of FD-PRO items. Statistics to evaluate the distribution were compiled – mean, SD, quartiles, percentage missing, percentage with the lowest possible score, and percentage with the best possible score. The skewness and kurtosis were computed as well.

Empirical Domain Specification and Scoring: Exploratory factor analysis (EFA) is a method for identifying the latent variables that give rise to the manifest variables (i.e., item responses). Factors can be interpreted as latent variables that account for common variance among items. Therefore, EFA was used to empirically determine the domain structure of the FD-PRO. The purpose of assessing domains was to evaluate the way that PRO items aggregated together to define relevant concepts. Inter-item correlations as well as modern psychometric methods (item response theory), were used to help determine the appropriate domain structure and scoring algorithm.¹³

Reliability: The reliability of FD-PRO scores was evaluated via internal consistency and test-retest reliability (TRTR).

Internal Consistency: Cronbach’s alpha was employed to evaluate internal consistency. Values of 0.70 and above indicated satisfactory internal consistency.¹³

Test-Retest Reliability: Test-retest reliability for FD-PRO scores was based on the two-way random intraclass correlation coefficient (ICC (2,1)) described by Shrout & Fleiss (1979).¹⁴ ICC(2,1) estimates of 0.70 and above indicated satisfactory retest reliability.^{14,15} Test-retest reliability was assessed in a group of patients who did not change significantly (stable retest sample). The absence of change in health status was operationalized by identifying patients for whom the anchor score did not change between week 1 and week 4. Both the PGA and the PGIS were used as anchors.

Validity: The validity of the instrument was assessed in terms of both concurrent validity and known-groups validity.

Concurrent validity: Criterion variables that were correlated with the FD-PRO are shown in Fig. 1. Correlations between the FD-PRO scores and continuous criterion variables were computed via Spearman correlations, whereas correlations with categorical criterion variables were computed via polyserial correlations. Concurrent validity was estimated

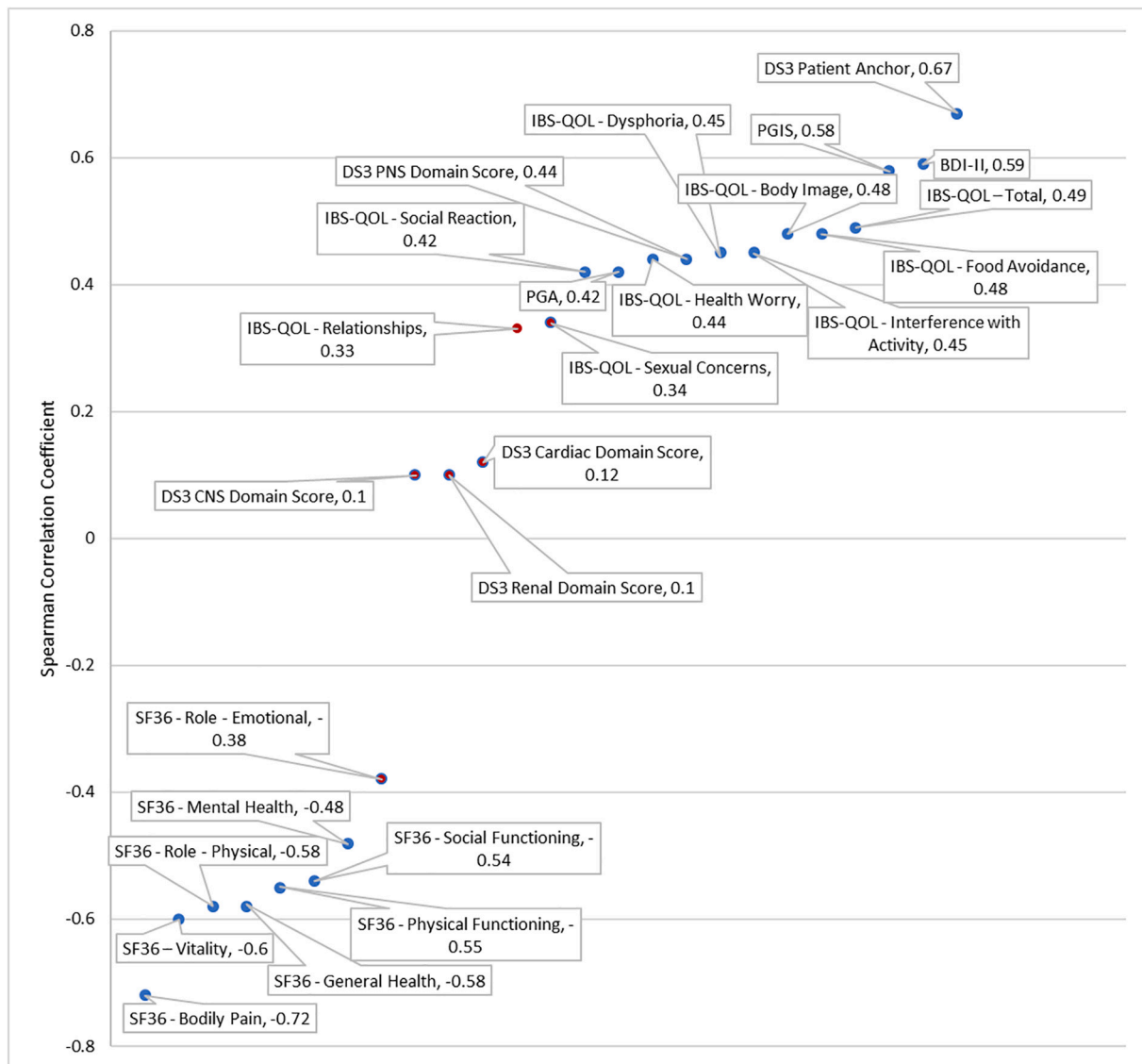


Fig. 1. Correlations between FD-PRO TSS and criterion variables.

BDI-II: Beck Depression Inventory Second Edition; CNS: central nervous system; DS3: Fabry Disease Scoring System; FD-PRO TSS: Fabry disease patient-reported outcome total symptom score; IBS-QoL: Irritable Bowel Syndrome-Quality of Life; PGA: Physician Global Assessment; PGIS: Patient Global Impression-Static; PNS: peripheral nervous system; SF-36: Medical Outcomes Study 36-Item Short Form Survey.

using Week 1 FD-PRO scores and variables collected at baseline. Estimated correlations greater than or equal to approximately $|0.4|$ were considered indicative of satisfactory concurrent validity.

Known-Groups Validity: The PGIS, PGA, DS3 patient assessment, and the symptom cohort were used in the known-groups validity analysis. Symptom cohort was determined at screening based on the responses to the screening questionnaire. Additionally, each clinical assessment in the DS3 was used to evaluate known-groups validity. Many of these assessments were based on biomarker and/or physiological data and thus were not expected to be as strongly related to the FD-PRO score as the other known-groups validators. The known-groups validity analysis entailed fitting a linear model (i.e., a regression model) with the FD-PRO scores as the dependent variable and the known-groups validator as the independent variable. A separate model was fit for each known-groups validator. In each analysis, the asymptomatic level of the known-groups validator was employed as the reference group. All other known health states assessed in the known-groups validator were contrasted against this reference group. The differences in known-groups was used to establish that the FD-PRO scores differed appropriately across known health groups, with statistical significance of differences and corresponding effect sizes (i.e., R^2) reported. A p -value of less than 0.05 was interpreted as evidence that the known-health group was associated with different FD-PRO scores than the reference group. The R^2 was evaluated as an overall measure of the degree of association between the known-groups validator and the FD-PRO score.

All subsequent analyses described were performed using R statistical software (R, Version 3.4.4, R Development Core Team).

3. Results

Patients: A total of 138 FD patients met eligibility criteria. Mean age was 43 years (range 18 to 72 years). There were slightly more males (52.9%) than females (47.1%), and the majority were White (84.1%). Classic phenotype represented most study participants (59.4%), followed by late-onset (25.4%) and unknown (15.2%) phenotypes. Almost three quarters (71.7%) of the patients were ERT-treated while the remaining (28.3%) were treatment naïve. About a third (31.9%) of patients had documented analgesic use. Regarding disease severity, the screening questionnaire revealed that patients with moderate disease comprised the largest group (Table 2).

Compliance and time to completion: Weekly compliance rates were high and consistent from Week 1 to Week 4, with approximately 87% of patients completing a minimum of 4 diaries each week (Table 3). Average time to completion of the electronic diary FD-PRO ranged from 2 min and 51 s for Week 1, to 1 min and 57 s for Week 4.

Item-response distributions: Weekly averages of item-responses from Weeks 1 to 4, shown in Table 4, revealed that item responses were skewed toward zero, indicating relatively modest symptom severity.

Empirical Domain Specification: Application of modern psychometric methods (IRT)¹⁶ yielded several important results: First, exploratory factor analysis (EFA)^{17,18} model fit statistics indicated that a two-factor solution was appropriate, accounting for 97% of the variance, with items 1 through 8 loading on the first factor, items 9 through 16 loading on the second factor, and item 17 (sweating) did not load on either factor. Because the EFA indicated that the sweating item was not related to the other FD-PRO items, it was dropped from scoring. Second, the local dependence statistics (G2 developed by Chen and Thissen)^{19–21} generated via IRT models, indicated that items 1–8 did not form a second distinct factor, but rather were essentially the same item repeated.²² That is, the local dependence statistics indicated that items 1–8 evaluating neuropathy all performed identically and thus would be most appropriately scored as a single item. This IRT evidence was supported by the high inter-item correlations (as high as 0.89 among the 8 items). This combination of evidence suggested that a one-factor model was more appropriate than a 2-factor model. To address the local

Table 2
Participant demographics and baseline clinical information.

Parameter	FD patient data (N = 138)
Age, years; Mean (SD)	43 (13.7)
Sex, n (%)	
Male	73 (52.9)
Female	65 (47.1)
Race, n (%)	
White	116 (84.1)
Asian	20 (14.5)
Multiracial	1 (0.7)
American Indian or Alaska native	1 (0.7)
Ethnicity, n (%)	
Hispanic or Latino	68 (49.3)
Not Hispanic or Latino	68 (49.3)
Not reported	2 (1.4)
Phenotype ^a , n (%)	
Classic	82 (59.4)
Late-Onset	35 (25.4)
Missing	21 (15.2)
Country, n (%)	
Argentina	20 (14.5)
Australia	18 (13)
Brazil	18 (13)
Canada	8 (5.8)
Czech Republic	18 (13)
Germany	1 (0.7)
Japan	8 (5.8)
Korea	2 (1.4)
Portugal	20 (14.5)
Taiwan	10 (7.2)
USA	15 (10.9)
Previous treatment, n (%)	
ERT-treated	99 (71.7)
Treatment naïve	39 (28.3)
Analgesic use, n (%)	44 (31.9)
Symptom severity at screening ^b , n (%)	
Absent	16 (11.6)
Mild	34 (24.6)
Moderate	55 (39.9)
Severe	3 (2.9)

ERT: enzyme replacement therapy; FD: Fabry Disease; SD: standard deviation.

^a Phenotype based on characteristic *GLA* mutation (both sex) or residual enzyme activity (males only).

^b Based on responses of 0 (absent), 1 to 3 (mild), 4 to 7 (moderate) and 8 to 10 (severe) on items in the screening questionnaire.

Table 3
Compliance and time to completion of FD-PRO instrument.

Week	Compliance, N (%) ^a	Completion, N (%) ^b	Time to completion (min: s)		
			Median	Mean	SD
Week 1	121 (87.7%)	115 (83.3%)	2:33	2:51	1:27
Week 2	121 (87.7%)	111 (80.4%)	1:52	2:13	1:26
Week 3	120 (87.0%)	106 (76.8%)	1:49	2:01	1:06
Week 4	120 (87.0%)	106 (76.8%)	1:40	1:57	1:13

FD-PRO: Fabry disease patient-reported outcome; SD: standard deviation.

^a Number of patients that reported a minimum of 4 daily diaries each week; ^b Number of patients in compliance and with valid timestamps for that week; these patients were used to compute the time to completion values. Valid timestamps were defined as diary entries where the time to completion was within the 0 to 30-min range, as pre-programmed in the device. Weeks 1–4 had 6, 10, 14, and 14 patients with not-valid timestamps, respectively.

dependence, a neuropathy parcel was created by computing the average of items 1–8 and rounding to the nearest integer. This parcel was treated as a single item that replaced items 1–8. Furthermore, the local dependence statistics also indicated that items 13, 15, and 16 evaluating audiovisual impairments all performed identically as well and would be appropriately scored as a single item. Therefore, an audiovisual parcel was created with items 13, 15, and 16. Finally, model fit statistics

Table 4
Weekly item response distribution

Item	Week 1 (N = 121)			Week 2 (N = 121)			Week 3 (N = 120)			Week 4 (N = 120)		
	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Pain in hands or arms	0.86	1.70	1.99	0.86	1.76	2.09	1.00	1.84	1.97	1.00	1.83	2.14
Burning feeling in hands or arms	0.57	1.27	1.74	0.43	1.31	1.89	0.71	1.41	1.74	0.57	1.41	1.88
Numbness in hands or arms	0.57	1.48	1.89	0.43	1.59	2.04	0.43	1.61	1.99	0.50	1.64	2.11
Tingling in hands or arms	0.50	1.28	1.72	0.43	1.36	1.85	0.40	1.45	1.87	0.29	1.41	1.96
Pain in feet or legs	1.00	2.01	2.32	1.14	2.14	2.43	1.43	2.18	2.34	1.33	2.24	2.45
Burning feeling in feet or legs	0.43	1.34	1.84	0.43	1.40	2.04	0.67	1.56	1.99	0.50	1.61	2.12
Numbness in feet or legs	0.20	1.33	1.86	0.43	1.50	2.07	0.33	1.61	2.14	0.29	1.67	2.19
Tingling in feet or legs	0.33	1.23	1.73	0.43	1.40	1.99	0.43	1.48	1.99	0.29	1.48	2.03
Headache	0.71	1.66	2.10	1.14	1.80	2.21	1.17	1.74	2.03	0.67	1.74	2.18
Abdominal pain	0.67	1.65	2.09	0.86	1.69	2.11	0.83	1.60	1.98	1.00	1.69	2.12
Heat intolerance	1.20	1.94	2.23	1.00	1.96	2.40	1.00	1.97	2.27	1.14	1.99	2.32
Swelling in lower extremities	0.43	1.30	1.80	0.40	1.51	2.09	0.71	1.60	2.06	0.40	1.55	2.12
Tinnitus	0.43	1.54	2.25	0.20	1.73	2.50	0.67	1.74	2.37	0.50	1.71	2.39
Tiredness/fatigue	3.00	3.33	2.65	2.86	3.34	2.79	3.14	3.36	2.71	2.86	3.27	2.67
Hearing impairment	0.14	1.19	2.00	0.00	1.31	2.24	0.00	1.25	2.11	0.00	1.29	2.22
Vision impairment	0.00	1.14	1.96	0.00	1.19	2.04	0.00	1.21	2.10	0.00	1.28	2.20
Sweating ^a	3.00	3.22	2.17	3.40	3.46	2.32	3.00	3.17	2.08	3.25	3.21	2.22
Sweating ^b	0.71	1.38	1.72	0.60	1.43	1.83	0.71	1.26	1.64	0.50	1.24	1.61

^a This item included a gatekeeper: patients were asked “in the past 24 hours, were you in a situation that should have led to sweating?” Only patients who responded “yes” were invited to respond to the Sweating item; ^b Item response distribution using a scoring approach attributing a score of ‘0’ if the response to the gatekeeper item was ‘No/Not sure’.

SD: standard deviation.

(RMSEA, TLI, CFI)²³ derived from a (one-factor) confirmatory factor analysis (CFA)²⁴ showed that the final 7 item scale comprising of the neuropathy parcel, the audiovisual parcel, and 5 additional items (headache, abdominal pain, heat intolerance, swelling in lower extremities, tiredness/fatigue) was unidimensional and thus a single total symptom score (TSS) was a suitable representation of the scale.

3.1. Scoring

The final scoring method was determined to be the following: the mean of the two item parcels and five individual items, averaged to week level. The algorithm can be written as:

$$\text{Integer} \left(\frac{\sum_{i=1}^8 \text{item}_i}{8} \right) + \sum_{i=9}^{12} \text{item}_i + \text{item}_{14} + \text{Integer} \left(\frac{\sum_{i=13}^3 \text{item}_{i5} \text{item}_{i6}}{3} \right)$$

where “Integer” indicates a function for rounding a number to the nearest integer.

The in weekly distribution of the FD-PRO TSS ranged from an average of 1.80 at Week 1 (median = 1.37, SD = 1.63) to 1.90 at Week 4 (median = 1.45, SD = 1.75).

3.2. Reliability

Internal consistency: The FD-PRO TSS had a very high degree of internal consistency, with Cronbach’s alpha ranging from 0.89 to 0.91 over the course of the 4 weeks.

Test-retest reliability: The ICC value for patients with the same PGIS anchor score at Week 1 and Week 4 was 0.96 ($n = 57$). Similarly, the ICC value for patients with the same PGA anchor score at Week 1 and Week 4 was 0.91 ($n = 83$). These values indicated a high degree of score reproducibility in the FD-PRO TSS.

Concurrent validity: Correlations between the FD-PRO TSS and the criterion variables were high, suggesting that the FD-PRO TSS captured diverse relevant symptoms experienced by FD patients (Fig. 1). The correlation between the SF-36 score and the FD-PRO TSS exceeded the ≥ 0.40 threshold for all SF-36 domains excepted for the Role – Emotional domain. Similarly, all IBS-QOL subdomains except two (sexual concerns, relationships) had correlations ≥ 0.40 . Strong correlations were also observed between the FD-PRO TSS and the DS3 patient anchor, the

PGIS, and the BDI-II scale. Correlation between the FD-PRO TSS and the clinician PGA was weaker, perhaps reflecting the discrepancy between PROs and ClinROs. DS3 domains that included biomarker/physiological data had low correlations with the FD-PRO TSS (e.g., Renal domain score correlation = 0.10).

Known-groups validity: The results of the known-groups validity analysis are presented in Table 5. The PGA was used to establish known health groups and a model was fit with those groups as the independent variables (predictor) with the FD-PRO TSS at week 1 as a dependent variable (outcome). The FD-PRO TSS associated with the PGA groups of “very mild symptoms” and “mild symptoms” was not different from that associated with the reference group of “no symptoms” ($P = 0.050050$ and $P = 0.075075$, respectively). However, the “moderate symptoms” and “severe symptoms” groups had significantly different FD-PRO TSS compared with the reference group ($P < 0.001001$ and $P < 0.001$, respectively). Similarly, the PGIS was used to establish known health groups and a model was fit with those groups as the independent variable with the FD-PRO TSS at week 1 as a dependent variable. The PGIS “mild symptoms” group did not have a different FD-PRO TSS compared to the “no symptoms” reference group ($P = 0.081081$) while “moderate symptoms” and “severe symptoms” groups were associated with significantly different TSS compared with the reference group ($P < 0.001001$ and $P < 0.001001$, respectively). The R^2 values of both the model using the PGA known-groups validator and the model using the PGIS known-groups validator were relatively high (0.20 and 0.33, respectively). Note that this difference in R^2 may be explained by the recall period: the PGA recall period was for the day the anchor was collected, rather than the week recall period employed in both the PGIS and FD-PRO TSS.

Similar to the PGA and PGIS, the symptom cohort at screening was used to establish known health groups and a model was fit with those groups as the dependent variable with the FD-PRO TSS at week 1 as a dependent variable. Using the symptom cohort at screening, the FD-PRO TSS in the group with mild symptoms was not significantly different than FD-PRO TSS in the reference group (i.e., patients with no symptoms) ($P = 0.542542$). In both moderate and severe cohorts, FD-PRO TSS was significantly higher compared to the reference group ($P = 0.042042$ and $P < 0.001$, respectively). The $R^2 = 0.38$ suggested a strong relationship between the FD-PRO TSS and the symptom cohorts. Finally, using the DS3 patient anchor, the FD-PRO TSS was statistically significantly higher in groups 2, 3, and 4 than the reference group 0 ($P \leq$

Table 5
Known Groups Validity Regression of FD-PRO TSS using PGA anchor, PGIS anchor, symptom severity at screening or DS3 patient anchor.

Instrument, Levels of severity	N	Group Average	Estimated Difference from Reference Group	95% CI	p-value
PGA anchor					
None - Reference Group		0.18	–	–	–
Very Mild	27	1.43	1.25	0.01-2.50	0.050050
Mild	36	1.28	1.10	-0.11-2.31	0.075075
Moderate	41	2.48	2.30	1.10-3.50	<0.001
Severe	9	3.06	2.88	1.40-4.36	<0.001
Very Severe	1	1.93	1.75	-1.39-4.89	0.272272
Model - R ²			0.20	–	–
PGIS anchor					
None - Reference Group	22	0.69	–	–	–
Mild	42	1.34	0.66	-0.08-1.39	0.081081
Moderate	34	2.85	2.17	1.40-2.93	<0.001
Severe	8	3.58	2.89	1.74-4.05	<0.001
Very Severe	0	–	–	–	–
Model - R ²			0.33	–	–
Symptom Severity at Screening					
Absent - Reference Group	11	0.68	–	–	–
Mild	32	0.96	0.28	-0.62-1.18	0.542542
Moderate	48	1.58	0.89	0.03-1.76	0.042042
Severe	30	3.44	2.76	1.85-3.67	<0.001
Model - R ²			0.38	–	–
DS3 Patient anchor					
0 (best overall well-being) - Reference Group	24	0.53	–	–	–
1	40	1.07	0.55	-0.09-1.18	0.090090
2	29	2.64	2.12	1.44-2.80	<0.001
3	21	3.40	2.88	2.14-3.61	<0.001
4 (worst overall well-being)	2	2.50	1.97	0.16-3.78	0.033033
Model - R ²			0.44	–	–

CI: confidence interval; DS3: Fabry Disease Scoring System; FD-PRO TSS: Fabry Disease Patient-Reported Outcome total symptom score; PGA: Physician Global Assessment; PGIS: Patient Global Impression-Static.

0.001001, $P \leq 0.001001$, and $P = 0.033033$, respectively). The value $R^2 = 0.44$ was the highest among all the anchors, reflecting a robust relationship between the FD-PRO TSS and the DS3 patient anchor.

A regression model was fit for each DS3 clinical assessment separately. Overall, the known-groups from the clinical assessments were not strongly associated with different week 1 FD-PRO TSS (results not presented). This was likely due to the fact that many of the clinical assessments were based on biomarker/physiological data. This created known-groups that were only modestly associated with the patient-reported FD-TSS.

4. Discussion

FD presents with a constellation of clinical signs and symptoms due to progressive accumulation of GL3 in organ systems and the heterogeneity of clinical manifestation across sex and mutation type. The early

symptoms, including chronic neuropathic pain and episodic severe pain crises, typically emerge during childhood, especially in classic disease. Symptoms such as hypohidrosis, gastrointestinal (GI) disturbances (bloating, diarrhea, abdominal pain), and signs of characteristic skin abnormalities (angiokeratomas) and asymptomatic corneal opacity (cornea verticillata) are additional common early manifestations. Other symptoms include auditory loss, tinnitus, and visual impairment due to involvement of the peripheral nervous system and fatigue.²⁵ Organ damage is typically due to renal (albuminuria, glomerulosclerosis and chronic kidney disease that progress to renal failure), cardiac (left ventricular hypertrophy [LVH] associated with myocardial fibrosis and arrhythmias), and cerebrovascular (transient ischemic attacks and strokes) involvement, leading to premature death. The late-onset phenotype presents with typical cardiac symptoms (e.g., LVH, arrhythmia) and, in some cases, decreased glomerular filtration rate (GFR) present in the fourth to seventh decades of life, reflecting delayed onset and slower disease progression.

While assessment of organ involvement based on laboratory, histological and imaging methods provide definitive evidence of the severity and progression of disease, PROs are useful tools to understand the disease burden, improvement, prevention of disease progression, and symptom burden. These are important objectives in FD treatment and management, particularly given the high variability in disease manifestation across patient populations. A recent European expert consensus statement on therapeutic goals in FD highlighted the importance of quality of life and long-term prognosis in the management of FD.²⁶ In a systematic review of literature conducted by Arends et al.,²⁷ the authors highlighted that the PRO instruments used in clinical studies in FD are generic and do not reflect the full spectrum and accurate evaluation of disease burden. Despite the significant symptomatic symptom burden in Fabry patients, the effect of ERTs on patients' experiences is inconclusive which may be attributed to the small size of the studies and the lack of a disease-specific PRO instrument.

To address this gap, the FD-PRO was developed in accordance with robust qualitative research guidelines^{4,28,29} to systematically measure symptom severity with items that assess neuropathic symptoms (pain, tingling, numbness and burning in upper/lower extremities), headache, abdominal pain, heat intolerance, swelling, tinnitus, fatigue, hearing/vision impairment, and hypohidrosis (diminished sweating) in the past 24 h. A final item that assesses difficulty engaging in regular physical activities in the past 24 h is a measure of exercise intolerance, a notable impact of FD on patients. The 24-h recall period was selected as symptoms may wax and wane over time; hence it is expected that the daily administration of the PRO over a week will be a better reflection of how a patient feels. The current study expands upon the qualitative research by testing the FD-PRO quantitatively to evaluate the measurement properties of the instrument.⁶

Our results show that an empirically derived total symptom score (TSS) can be calculated from the FD-PRO to represent the composite symptom severity as a weekly average. The TSS exhibits acceptable measurement properties as demonstrated by: (1) the reliability of the TSS measure exceeded well-established criteria for both internal consistency and repeatable reliability (test-retest), (2) evidence of concurrent validity as the TSS was highly correlated ($r \geq |0.4|$) with well-established and conceptually similar PROs and ClinROs (e.g., physical symptoms evaluated by the generic SF-36 instrument, the IBS-QOL, and the BDI-II), (3) evidence of discriminant validity as the TSS was not highly correlated ($r < |0.4|$) to constructs that are theoretically different (e.g., DS renal, cardiac, CNS scores), and (4) evidence of known-groups validity as the TSS discriminated moderate/severe from least severe FD groups (based on PGIS, PGA, baseline symptom severity, and DS3 patient anchors) in known-groups validity analyses. Furthermore, high compliance rates were observed for the FD-PRO instrument and the time required for patients to complete the daily electronic diary was minimal, taking less than 3 min on average in Week 1 and less than 2 min in Week 4. Both the electronic nature of the instrument and the short time to

complete makes the FD-PRO a user-friendly tool to assess symptoms of patients with FD.

Evaluating meaningful change and sensitivity to change was not feasible in this study. The number of patients reporting change in symptom severity was low, which was consistent with 72% of the patients having stable symptoms at screening (Table 2). A forthcoming study is being conducted over a longer follow up period and results from this study will be used to demonstrate sensitivity to change and meaningful change for the FD-PRO. The forthcoming study will provide an evaluation of the therapeutic effects of currently available therapies over a longer period of time. As some domains may be more sensitive to detect clinically meaningful change, identification of domains that are responsive to change will be further evaluated in the forthcoming study.

This study has important implications in FD for the following reasons. The validation of the FD-PRO in this study addresses an unmet need as this is the first disease-specific PRO in FD that has been tested psychometrically. The large number of participants in this study enabled robust psychometric testing, a noteworthy highlight in such a rare disease. One key limitation is that the design of the study, which included mostly stable patients with limited changes in symptoms during the one-month interval, precluded the determination of a threshold for meaningful within-patient change on the FD-PRO TSS. Patients reported only modest symptoms at baseline which explained the skewness of responses toward 0 in the item response analysis. Future studies of FD patient populations who are actively symptomatic and followed up for a longer duration will be required for testing the FD-PRO on its ability to capture the natural history of the disease and the effect of treatment. Moreover, further evidence is needed to evaluate if the sensitivity to change and response to treatment is captured by the TSS alone or if addition item/domain level analyses is warranted. For example, neuropathy and abdominal symptoms have been postulated to have significant impact on the quality of life of Fabry patients and may be more likely to respond to treatment. Nonetheless, our study is an important step in advancing the development of the novel FD-PRO.

Psychometric and statistical evidence indicate that the FD-PRO is a reliable, valid, user-friendly and robust measure of FD symptom severity, with minimal burden of administration. The instrument is appropriate to measure the overall symptom severity and applicable for use in clinical studies.

Funding

This study was funded by Sanofi Genzyme.

Authors' contributions

AH, PD, CG, CI, and DS contributed to the study design, analysis, and interpretation of data. NL, VM, and JP contributed to the interpretation of the data. All authors were involved in the interpretation of the data, drafting, and critical revision of the manuscript. All authors are accountable for the accuracy and integrity of the manuscript.

Declaration of Competing Interest

AH, PD, and NL are employees of Sanofi Genzyme; VM was an employee of Sanofi Genzyme at the time this study was conducted. AH, PD, and NL hold stock of Sanofi Genzyme. CI and DS are employees of Pharmerit, a research consulting firm, and hold no Sanofi Genzyme stock. CG is the owner of Gwaltney Consulting, a research consulting firm, and holds no Sanofi Genzyme stock. CI, DS, and CG were paid consultants of Sanofi Genzyme. JP is a consultant for Sanofi Genzyme and receives speaker fees from Sanofi Genzyme.

Acknowledgments

The authors would like to thank all of the patients who participated

in this study. We also thank Frederic Le-Foll and Laurence Salin for their contributions to the conduct of the study.

References

- [1] R. El-Abassi, D. Singhal, J.D. England, Fabry's disease, *J. Neurol. Sci.* 344 (1–2) (2014) 5–19.
- [2] D.P. Germain, Fabry disease, *Orphanet J. Rare Dis.* 5 (2010) 30.
- [3] R. Schiffmann, D.G. Warnock, M. Banikazemi, et al., Fabry disease: progression of nephropathy, and prevalence of cardiac and cerebrovascular events before enzyme replacement therapy, *Nephrol. Dial. Transplant.* 24 (7) (2009) 2102–2111.
- [4] C. Simoncini, L. Chico, D. Concolino, et al., Mitochondrial DNA haplogroups may influence fabry disease phenotype, *Neurosci. Lett.* 629 (2016) 58–61.
- [5] US Food and Drug Administration, Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, 2009.
- [6] A. Hamed, P. DasMahapatra, N. Lyn, et al., Development of the fabry disease patient-reported outcome (FD-PRO): a new instrument to measure the symptoms and impacts of fabry disease, *Orphanet J. Rare Dis.* 16 (2021) 285.
- [7] Good epidemiology practice (GPP). Guidelines for Good Pharmacoeconomics Practice, 2020. Last accessed 22 October, <https://www.pharmacoepi.org/resources/policies/guidelines-08027/>.
- [8] IEA guidelines for proper conduct of epidemiological research – HRSA, 2020. Last accessed 22 October, <https://www.yumpu.com/en/document/view/19409945/iea-guidelines-for-proper-conduct-of-epidemiological-research-hrsa>.
- [9] International Fabry disease phenotype-genotype database, 2020. Last accessed 22 October, <http://dbfpg.org/dbfpg/fabry/>.
- [10] J.E. Ware, M.R. Kosinski, J.B. Bjorner, D.M. Turner-Bowker, B. Gandek, M. E. Maruish, User's Guide for the SF-36v2 Health Survey, QualityMetric Incorporated, Lincoln, R.I., 2007.
- [11] D.L. Patrick, D.A. Drossman, I.O. Frederick, J. DiCesare, K.L. Puder, Quality of life in persons with irritable bowel syndrome: development and validation of a new measure, *Dig. Dis. Sci.* 43 (2) (1998) 400–411.
- [12] S.J. Lewis, K.W. Heaton, Stool form scale as a useful guide to intestinal transit time, *Scand. J. Gastroenterol.* 32 (9) (1997) 920–924.
- [13] L.B. Mokkink, Prinsen CAC, D.L. Patrick, COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs): user manual 78. Amsterdam, the Netherlands: COSMIN, 2018.
- [14] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, *Psychol. Bull.* 86 (2) (1979) 420–428.
- [15] C.J. Gwaltney, A.L. Shields, S. Shiffman, Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review, *Value Health* 11 (2008) 322–333.
- [16] R.P. McDonald, Test Theory: A Unified Treatment, 1st ed, Psychology Press, 1999.
- [17] K.J. Preacher, R.C. MacCallum, Repairing Tom Swift's electric factor analysis machine, *Understanding statistics: Statistical issues in psychology, education, and the social sciences.* 2 (1) (2003) 13–43.
- [18] C.A. Bernardis, R.I. Jennrich, Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis, *Educ. Psychol. Meas.* 65 (2005) 676–696.
- [19] W.-H. Chen, D. Thissen, Local dependence indexes for item pairs using item response theory, *J. Educ. Behav. Stat.* 22 (3) (1997) 265–289.
- [20] Y. Liu, A. Maydeu-Olivares, Local dependence diagnostics in IRT modeling of binary data, *Measurement* 73 (2) (2012) 254–274.
- [21] C.R. Houts, M.C. Edwards, The performance of local dependence measures with psychological data, *Appl. Psychol. Meas.* 37 (7) (2013) 541–562.
- [22] E.H. Ip, Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models, *Br. J. Math. Stat. Psychol.* 63 (2) (2010) 395–416.
- [23] L.T. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Struct. Equ. Model. Multidiscip. J.* 6 (1) (1999) 1–55.
- [24] K.A. Bollen, Structural Equations with Latent Variables. 1st Edition ed, 1989.
- [25] A. Ortiz, D.P. Germain, R.J. Desnick, J. Politei, M. Mauer, A. Burlina, C. Eng, R. J. Hopkin, D. Laney, A. Linhart, S. Waldek, E. Wallace, F. Weidemann, W. R. Wilcox, Fabry disease revisited: management and treatment recommendations for adult patients, *Mol. Genet. Metab.* 123 (2018) 416–427.
- [26] C. Wanner, M. Arad, R. Baron, et al., European expert consensus statement on therapeutic goals in fabry disease, *Mol. Genet. Metab.* 124 (3) (2018) 189–203.
- [27] M. Arends, C.E.M. Hollack, M. Biegstraaten, Quality of life in patients with fabry disease: a systematic review of the literature, *Orphanet J. Rare Dis.* 10 (2015) 77.
- [28] D.L. Patrick, L.B. Burke, C.J. Gwaltney, N.K. Leidy, M.L. Martin, E. Molsen, L. Ring, Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument, *Value Health* 14 (2011) 967–977.
- [29] D.L. Patrick, L.B. Burke, C.J. Gwaltney, N.K. Leidy, M.L. Martin, E. Molsen, L. Ring, Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 2—eliciting concepts for a new PRO instrument, *Value Health* 14 (2011) 978–988.