


# Development and Preliminary Psychometric Evaluation of the Glycogen Storage Disease Type Ia Functional Assessment Diary (GSD FAD)

Diane M Turner-Bowker<sup>1</sup>, Shayna Egan<sup>1</sup>, Jessica Butler<sup>1</sup>, Adrian Jewett<sup>2</sup>, Deepali Mitragotri<sup>3</sup>, Brandon Foster<sup>2</sup>, Alison Skrinar<sup>1</sup>, Christina Theodore-Oklotka<sup>1</sup>

<sup>1</sup>Endpoint Development and Strategy, Ultragenyx Pharmaceutical Inc., Novato, CA, USA; <sup>2</sup>Patient Centered Outcomes, Lumanity, Long Beach, CA, USA; <sup>3</sup>Biostatistics & Epidemiology, Ultragenyx Pharmaceutical Inc., Novato, CA, USA

Correspondence: Diane M Turner-Bowker, Endpoint Development and Strategy, Ultragenyx Pharmaceutical Inc., 60 Leveroni Court, Novato, CA, 94949, USA, Tel +1 401 372 4793, Email DTurner-Bowker@ultragenyx.com

**Purpose:** Glycogen storage disease type Ia (GSDIa) is a rare, inherited, autosomal recessive condition with deficiency of glucose-6-phosphatase (G6Pase) characterized by fasting hypoglycemia due to an inability to release glucose from hepatic glycogen stores and other metabolic pathways, requiring frequent consumption of exogenous glucose for survival. Patient-reported outcome (PRO) measures are important to assess GSDIa burden, though no disease-specific PRO for GSDIa exists. This research describes the Glycogen Storage Disease Functional Assessment Diary (GSD FAD), a 31-item PRO developed to assess the signs/symptoms and impacts of GSDIa in individuals ages 8 and older.

**Patients and Methods:** Mixed methods research including a literature review, online survey, and concept elicitation (CE) interviews informed the construction of the GSD FAD. Cognitive debriefing (CD) interviews evaluated concept relevance and understanding of the questionnaire. Psychometric properties were evaluated using Phase 3 trial screening/baseline data to assess item variability, structure/scaling potential, scoring, reliability, validity, and minimum detectable change.

**Results:** The literature review, online survey (N=26), and CE interviews (N=7) identified hypoglycemia and cornstarch regimen impacts as most burdensome and most important to treat from the patient perspective. The initial draft GSD FAD included 36 items. Most items were interpreted as intended in CD interviews (N=16); revisions to the GSD FAD addressed interpretation issues. Three domain scores (Symptoms Total Score, Sleep Impacts Total Score, and Daily Impacts Total Score) showed acceptable reliability and validity but had notable ceiling effects that may limit responsiveness in some settings.

**Conclusion:** The GSD FAD is a novel, content-valid PRO that measures the humanistic burden of GSDIa and cornstarch treatment regimen, including the signs/symptoms of hypoglycemia and their impact on health-related quality of life (HRQoL) in individuals  $\geq 8$  years with GSDIa, that yields reliable and valid scores. While developed in a trial setting, the GSD FAD has potential for use in nutritional, behavioral, and/or educational applications.

**Keywords:** GSDIa, patient-reported outcome, content validity, psychometric, reliability, validity

## Introduction

Glycogen storage disease type Ia (GSDIa) is a rare, autosomal recessive disease with a deficiency of glucose-6-phosphatase (G6Pase), an enzyme essential for glycogenolysis and gluconeogenesis.<sup>1,2</sup> G6Pase deficiency prevents glycogen breakdown, causing excess accumulation of glycogen in the liver and kidneys and life-threatening fasting hypoglycemia.<sup>3</sup> Additional metabolic abnormalities include hypertriglyceridemia, hyperuricemia, and lactic acidemia.<sup>2</sup>

There is no existing cure for GSDIa. It is managed with a medically-prescribed diet that typically includes frequent ingestion of large quantities of uncooked cornstarch/Glycosade<sup>®</sup> during the day and night to provide a steady supply of exogenous glucose.<sup>3</sup> Early treatment and close monitoring of glycemic control is essential; if patients with GSDIa have insufficient intake of carbohydrates, fatal hypoglycemia can result.<sup>4</sup> Patients living with GSDIa experience significant

burden from the condition, as well as its treatment and monitoring. Recent qualitative narrative research provided insight into the humanistic burden of GSDIa and its treatment, symptom burden, and diminished health-related quality of life (HRQoL) impacts on daily activities, emotional, social, relationships, and work/school function.<sup>5</sup>

Patients are experts in the experience of their disease and patient experience data (PED) are important in drug development to inform clinical trial design, endpoint selection, results interpretation, and benefit-risk assessment; health technology assessment in value assessments and cost-effectiveness analyses; disease management for population-based monitoring, case finding, benchmarking, and performance evaluation; and clinical practice for individualized care planning and patient-provider shared decision-making, among other applications.

Regulatory guidelines describe research methods to elicit key measurement concepts (signs/ symptoms and impacts) that are important and relevant to patients, as well as methods to evaluate patient-reported outcome (PRO) measures for use in assessing these concepts.<sup>6</sup>

Patient experience research is needed to understand GSDIa from the patient perspective and to inform measurement strategies for the assessment of health outcomes meaningful to patients with the condition. However, GSDIa patient-centered research following the methods outlined in regulatory guidelines is limited and to our knowledge, no GSDIa-specific PRO currently exists.

Generic PROs have been used in GSDIa,<sup>7–9</sup> however, items in generic PROs do not measure concepts that are relevant to individuals with GSDIa such as hypoglycemia events, symptoms of hypoglycemia (eg, tiredness, difficulty concentrating), and impacts specific to GSDIa (eg, constant monitoring of glucose levels, adjustment of cornstarch dosing, and strict lifelong dietary management and restrictions, fear of missing a dose with potentially fatal consequences, disrupted sleep due to nighttime cornstarch treatment, and limitations on physical, social, relationships and other aspects of daily life).<sup>10</sup> In following, generic PRO scores used in GSDIa may not be as sensitive as scores yielded from a disease-specific measure.

Comprehensive patient-based assessment includes both generic and disease-specific measures of health status. Disease-specific measures tend to yield sensitive scores that are more responsive to disease-specific changes, while generic measures make it possible to compare disease burden and outcomes across diseases and treatments.<sup>11</sup>

The current research identified the important and relevant signs/symptoms and impacts of GSDIa in individuals ages 8 and older and informed the development of a new disease-specific PRO, the Glycogen Storage Disease Functional Assessment Diary (GSD FAD). This novel 31-item PRO was developed following regulatory and scientific best practices<sup>6,12–15</sup> and is expected to fill specific, conceptual gaps with the assessment of clinically salient, patient-relevant items.

This manuscript describes the development of the GSD FAD, its content validation, and preliminary evidence regarding score psychometric performance using baseline data from a Phase 3, randomized, double-blind, placebo-controlled clinical trial to evaluate DTX401, an investigational adeno-associated virus serotype 8 (AAV8) vector containing the human *G6PC* gene (DTX401-CL301; ClinicalTrials.gov Identifier: NCT05139316) in pediatrics (8 to <18 years) and adults ( $\geq 18$  years) with GSDIa.

## Methods

### Development of the GSD FAD

#### Literature Review

A systematic, targeted concept-focused literature review was conducted to identify published literature describing signs/symptoms and impacts experienced by individuals living with GSDIa. A MEDLINE<sup>®</sup> search was conducted via PubMed, filtered to limit the search to relevant articles published between 2010 to 2020, in English language, and involving human participants. Key inclusion criteria for article selection included articles describing clinician- or patient-report of GSDIa signs/symptoms or impacts, or articles focused on PRO instruments developed using clinician or patient input to measure signs/symptoms and impacts of GSDIa. Key exclusion criteria excluded articles that were not focused on the signs/symptoms and impacts of GSDIa (eg, articles describing physiological measures of GSDIa, such as medical imaging).

Detailed search terms and criteria for full-text articles selection are described in [Supplementary Tables 1](#) and [2](#), respectively.

### Online Cross-Sectional Observational Survey

A cross-sectional, observational survey was conducted to assess patient experiences of GSDIa and, when appropriate, their caregivers.<sup>16</sup> Survey questions focused on symptoms and impacts of the condition, diet/eating, treatment regimen, and aspects of GSDIa to improve. Following WCG (Western) Institutional Review Board (IRB) ethics approval, participants were recruited through clinician referral, advocacy groups, and/or social media using an IRB-approved recruitment flyer. Eligible individuals completed the survey online on the SurveyMonkey platform accessed via a password-protected link.

### Concept Elicitation (CE) Interviews

Qualitative CE interviews were conducted to identify the important and relevant signs/symptoms and impacts of GSDIa from the perspective of patients and, when appropriate, caregivers. Following WCG IRB ethics approval, individuals were recruited from previous participation in the online cross-sectional observational survey, and through clinician referral, advocacy groups, and/or social media using an ethics-approved recruitment flyer. Face-to-face or telephone interviews were conducted using a semi-structured interview guide, lasted approximately 90 minutes, and were audio-recorded with consent. Recordings were transcribed, coded using Atlas.ti software package (Atlas.ti GmbH, Berlin) and qualitatively analyzed.<sup>17–19</sup>

### Questionnaire Construction

Converging results from the literature review, online survey, and patient interviews informed a preliminary conceptual model for GSDIa and development of the draft GSD FAD. The GSD FAD included a morning and evening diary intended for electronic administration in a clinical trial. Items were written following best practice principles for item construction<sup>20</sup> and most measure concept frequency using a 5-point verbal response scale (VRS), though a few items use additional response options.

### Combined CE and Cognitive Debriefing (CD) Interviews

Combined CE/CD interviews were conducted to confirm the GSD FAD content relevance in the target population, evaluate patient understanding of the content, and inform revisions to the draft measure. Participants were recruited in the same manner as for CE interviews. Interviews were approximately 90 minutes in duration, conducted via telephone or secure video chat that was audio recorded with consent, using a semi-structured interview guide.<sup>21</sup> Recordings were transcribed and content analyzed using Atlas.ti software package (Atlas.ti GmbH, Berlin).

Concept saturation was not formally evaluated due to the sample size; however, cognitive debriefing findings relating to content relevance were used to support the content validity evaluation.

### Questionnaire Revision

The GSD FAD was revised based on the CE/CD interview results and administered as an electronic daily diary in the Phase 3 DTX401-CL301 trial.

## Preliminary Psychometric Evaluation of the GSD FAD

Psychometric properties were evaluated using DTX401-CL301 Phase 3 screening/baseline data from the Intent-to-Treat (ITT) population (N=46), defined as all randomized participants who received any investigational product (DTX401 or Placebo). The GSD FAD was completed electronically by participants every morning and evening for 14 days prior to baseline and for 7 days prior to each applicable post-baseline visit. Trial data were analyzed for item variability, structure and scaling potential, scoring approach, score reliability, score validity, and preliminary score interpretation thresholds. For most analyses, listwise deletion was used to handle missing scores, except for the assessment of inter-item correlations, convergent validity, and discriminant validity, which used pairwise deletion. All analyses were conducted using R version 4.1.0<sup>22</sup> or higher, using a combination of base features and the following packages: tidyverse,<sup>23</sup> psych,<sup>24</sup> and irr.<sup>25</sup>

## Assessment of Item Variability

Item frequency distributions for daily data, and descriptive statistics for daily data aggregated over participants and days, were calculated to assess item completion and variability.

### Item Completion Rates

To examine completion patterns, missingness was calculated using the total number of participants at each day of the last 14 days of screening (Days -14 to -1) as the denominator and the total number of non-missing responses for each item as the numerator.

### Item Response Distributions

The frequency and percentage of responses on each category were calculated using the last 14 days of screening (Days -14 to -1) data for items to assess response category utilization and evaluate the presence of ceiling and floor effects (identified if the percentage of responses was highest in the response category representing the best possible health state [ie, ceiling effect] or the worst possible health state [ie, floor effect]).

### Assessment of Item Means and Modal Responses

Item response mean and mode were calculated for the last 14 days of screening (Days -14 to -1).

## Assessment of Structure and Scaling Potential

### Inter-Item Correlations

Spearman correlation coefficients ( $r_s$ ) were calculated between items within each of the last seven days of screening (Days -7 to -1) and were averaged across seven days to provide estimates of the inter-item associations in the presence of daily fluctuations. Correlation performance criteria were  $0.30 \leq |r_s| \leq 0.90$ .<sup>26</sup> Items showing poor performance were dropped from their respective domains, and correlations were re-examined.

### Item-Total Correlations and Internal Consistency Reliability

Classical test theory (CTT) methods were used to evaluate the scaling potential for hypothesized GSD FAD domains (symptoms, sleep impacts, daily impacts). Item discrimination was examined using corrected Pearson item-total correlations ( $r_p$ ) between each item and the sum score of the other items in the domain. A cutoff of  $r_p \geq 0.40$  was used to determine acceptable discrimination.<sup>27</sup> Overall internal consistency reliability was assessed with Cronbach's alpha ( $\alpha$ ) using data from the last seven days of screening (Days -7 to -1) averaged across seven days to provide more robust estimates of item/scale performance. A threshold of  $\alpha \geq 0.70$  was used to determine acceptable reliability,<sup>28-30</sup> and  $\alpha$  was iteratively reassessed with individual items removed.

## Scoring Approach

### Stability of Daily Scores

To evaluate the appropriate period over which to aggregate the daily scores, the stability of GSD FAD scores was assessed over a 7-day period using a two-way mixed effects intraclass correlation coefficient for the absolute agreement of average scores (ICC [A,K]).<sup>31</sup> An ICC (A,K) value of  $\geq 0.70$  was the acceptable threshold for the 7-day average score to be considered reliable.<sup>31-33</sup> The stability of daily scores was evaluated using data from the first and last 7-day periods of screening (Days -14 to -8 and Days -7 to -1). Additionally, the Spearman-Brown prediction formula (ICC[A,K] values from last seven days of screening [Days -7 to -1]) was used to examine the impact of missing daily scores on the reliability of 7-day average scores. A missing data rule requiring at least 4 of 7 observations to determine whether a participant's summary score was created or set to missing was applied. Recent research suggests that this commonly used rule was supported through simulations aimed to assess the impact of missing data on summary score estimates.<sup>34</sup> Specifically, this analysis evaluated the drop in reliability of the 7-day average score for up to 3 missing daily scores (ie, up to 50% missing). The 7-day average score was considered reliable if the estimate derived with less than seven days of data was  $\geq 0.70$ .

## Descriptive Summaries of Scores

Descriptive statistics for continuous variables were calculated, including sample size (N), mean, standard deviation (SD), median, interquartile ranges, minimum (Min), and maximum (Max). Score level floor and ceiling effects were evaluated. Ceiling effects were identified if the absolute difference between the minimum observed value and the mean divided by the SD was less than 1 (ie,  $[(\text{Min}-\text{Mean})/\text{SD}] < 1$ ). Floor effects were identified if the absolute difference between the maximum observed value and the mean divided by the standard deviation was less than 1 (ie,  $[(\text{Max}-\text{Mean})/\text{SD}] < 1$ ).<sup>35</sup> A value below 1 for either estimate indicated that scores were truncated for a substantial proportion of participants. Descriptive statistics were calculated for scores from the first and last 7-day periods of available screening data (ie, Days -14 to -8 and Days -7 to -1).

## Reliability and Validity

### Test-Retest Reliability

The test-retest reliability of 7-day average scores was examined across a 7-day lag using two screening time points, Days -14 to -8 and Days -7 to -1, using a two-way mixed effects intraclass correlation coefficient for the absolute agreement of single scores ICC(A,1). The criterion for acceptable reliability was an  $\text{ICC}(A,1) \geq 0.70$ .<sup>31-33</sup>

### Convergent and Discriminant Validity

Convergent and discriminant validity analyses examined the degree to which the Symptoms Total Score, Sleep Impacts Total Score, and Daily Impacts Total Score correlated as predicted with scores from other criterion measures (“co-validators”).<sup>27</sup> Spearman correlations ( $r_s$ ) using pairwise deletion examined the association between each GSD FAD domain score obtained from screening (Days -7 to -1) and scores from co-validators, including: Patient Global Impression of Severity (PGIS) scores (screening, Day -1), EQ-5D-(5L/Y) scores (baseline, Day 0), and starch diary scores (average frequency of nighttime cornstarch/Glycosade intake, average frequency of daytime cornstarch/Glycosade intake) (screening, Days -14 to -1). Correlations were characterized as  $|r_s| < 0.10$  (Discriminant),  $0.10 \leq |r_s| < 0.30$  (Small),  $0.30 \leq |r_s| < 0.50$  (Medium), and  $|r_s| \geq 0.50$  (Large).<sup>36,37</sup>

### Known-Groups Validity

Tests of known-groups assessed whether GSD FAD scores distinguished between groups known to differ clinically. The Symptoms Total Score was assessed using: Frequency of Hypoglycemic Events in Past Year ( $< 70$  mg/dl), Frequency of Hypoglycemic Events in Past Year ( $< 54$  mg/dl), PGIS-Had Low Blood Sugar, and PGIS-Severity of Symptoms. The Sleep Impacts Total Score was assessed using: Average Frequency of Daytime Cornstarch/Glycosade Intake, Average Total Daily Cornstarch/Glycosade Intake, Number of Hours of Fasting Tolerance, and PGIS – Bothered by Impacts. The Daily Impacts Total Score was assessed using: Average Frequency of Daytime Cornstarch/Glycosade Intake, Average Total Daily Cornstarch/Glycosade Intake, EQ-5D-(5L/Y) Usual Activities, and PGIS – Bothered by Impacts. Descriptive summaries of GSD FAD scores were calculated within known-groups, using GSD FAD domain scores obtained from screening (Days -7 to -1) against co-validating known-groups collected during screening or the baseline visit. The success criterion was monotonic ordering (either decreasing or increasing values) of a score from the GSD FAD score median across corresponding groups representing higher or lower levels of the measured construct.<sup>38</sup> Focus was placed on the median due to its resilience to outliers and skewed data.

### Minimum Detectable Change (MDC)

Distribution-based methods were used to calculate MDC indices (standard error of measurement [SEM], standard error of estimation [SEE], standard error of prediction [SEP], and reliable change index [RCI]) to establish criteria for identifying significant within-patient change for multi-item GSD FAD scores.<sup>39-42</sup> Each MDC index was multiplied by 1.96 to approximate a two-tailed significance test, establishing thresholds for significant within-patient change. The lowest and highest thresholds were used to define a range of significant within-patient change for each GSD FAD score. The inputs used to calculate these indices were the reliability estimates from the stability of daily scores analysis ( $\text{ICC}[A, K]$ ) and SDs for each GSD FAD score from the last 7-day period of the screening data (ie, Days -7 to -1).

## Results

### Development of the GSD FAD

Figure 1 presents the preliminary conceptual model for GSDIa based on converging results from the literature review, online survey, and CE interviews.

#### Literature Review

Seven articles were included in the final full-text review and used for contribution to the development of the conceptual framework. Sixty-three clinician- or patient-reported sign/symptom concepts and eight unique clinician- or patient-reported impacts of GSDIa were identified and used to inform the development of the online cross-sectional survey (Figure 1).

#### Online Cross-Sectional Observational Survey

Twenty-six participants from the United States (n=11 adults, n=6 adolescents, n=9 caregivers of children living with GSDIa) completed the online survey between September 2018 and March 2019 (Table 1). Respondents most frequently reported hypoglycemia (adults n=11, 100%; caregivers n=9, 100% [adolescents were not asked sign/symptom questions]). Common hypoglycemia triggers included increased physical activity (n=7 caregivers, 78%) and forgetting to consume cornstarch (n=10 adults, 91%). Frequently reported hypoglycemia symptoms were sweating (n=19, 95%), fatigue (n=17, 85%), hunger (n=16, 80%), shakiness (n=16, 80%), pale skin (n=13, 65%), and anxiety (n=8, 40%). Participants reported negative impacts of GSDIa on appearance (eg, weight), physical activities, daily activities (eg, chores, errands, school, work), sleep, emotional (eg, feeling self-conscious, embarrassed, frustrated, anxious, angry), relationships (eg, dependence on others), and social life (not being able to do things with friends) (Figure 1). Participants most frequently identified treatment regimen as the most bothersome aspect of GSDIa, and hypoglycemic events as the most important to improve.

#### Concept Elicitation (CE) Interviews

Seven participants (n=2 adults, n=3 adolescents, n=2 child and caregiver dyad) completed CE interviews (Table 2). Hypoglycemia was reported by all participants (n=7, 100%). Of the 29 signs/symptoms identified, the most frequently

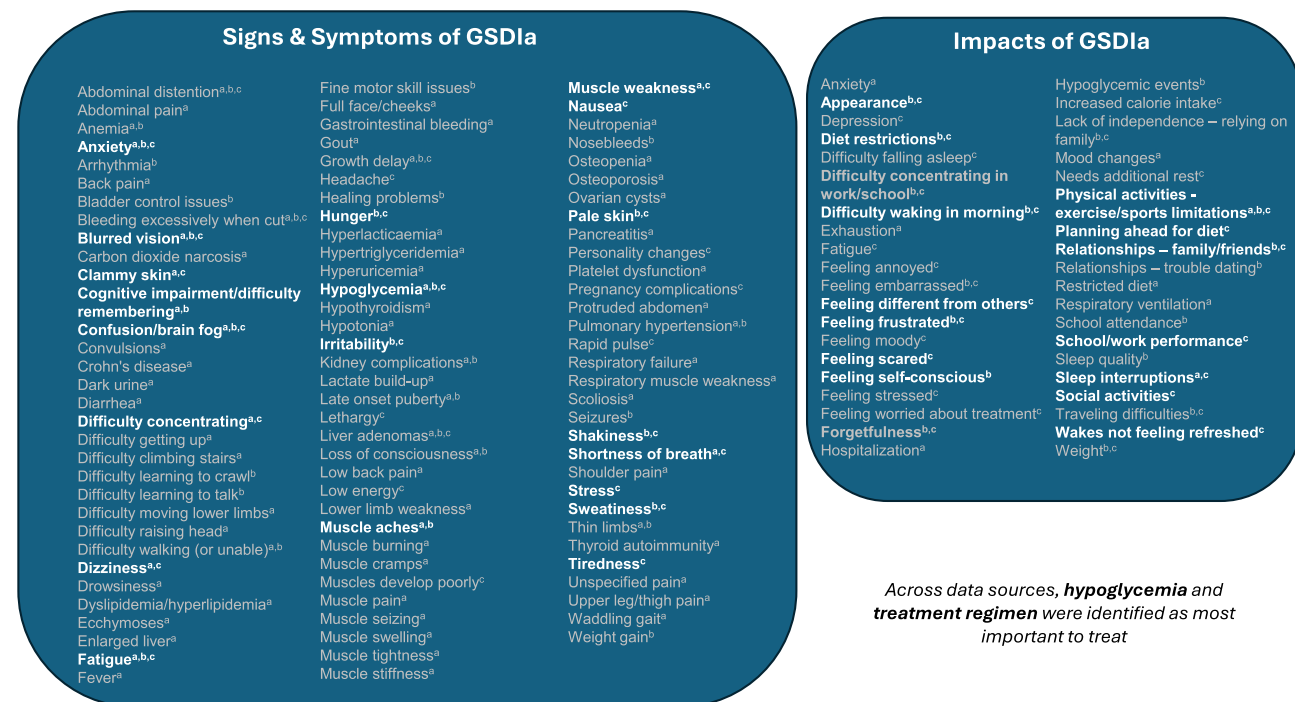


Figure 1 Reported in: <sup>a</sup>Literature review, <sup>b</sup>Online survey, <sup>c</sup>Concept elicitation (CE) interviews. Predominant concepts that appear in the final GSD FAD are bolded.

**Table 1** Participant Characteristics for the Online Survey (N=26)

Characteristic <sup>a</sup>	Caregivers (n=9) <sup>b</sup>	Adolescent Participants (n=9) <sup>b</sup>	Adolescents Who Completed a Survey (n=6) <sup>b</sup>	Adults (n=11)
Age (years)				
Min–Max	32–66	9–17	12–17	18–55
Mean (SD)	43.9 (10.2)	13.4 (3.2)	15.3 (2.0)	35.5 (10.5)
Gender (n, %)				
Female	6 (66.7%)	5 (55.6%)	1 (16.7%)	7 (63.6%)
Male	3 (33.3%)	4 (44.4%)	5 (83.3%)	4 (36.4%)

**Notes:** <sup>a</sup>Race/ethnicity data were not collected in the online survey. <sup>b</sup>Caregivers reported on behalf of children, adolescents, and themselves.  
**Abbreviation:** N, total of participants.

**Table 2** Participant Characteristics for the Concept Elicitation Interviews (N=7)<sup>a</sup>

Characteristic	Caregiver (n=1)	Child (n=1)	Adolescents (n=3)	Adults (n=2)
Age (years)				
Min–Max	N/A	N/A	16–17	20–46
Mean (SD)	45 (N/A)	8 (N/A)	17 (0.6)	33 (18.4)
Gender (n, %)				
Female	0 (0.0%)	1 (100.0%)	2 (66.7%)	1 (50.0%)
Male	1 (100.0%)	0 (0.0%)	1 (33.3%)	1 (50.0%)
Ethnicity (n, %)				
Hispanic/Latino	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Non-Hispanic/Latino	1 (100.0%)	1 (100.0%)	3 (100.0%)	2 (100.0%)
Race (n, %)				
American Indian or Alaska Native	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Asian	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Black or African American	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Native Hawaiian or Other Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
White	1 (100.0%)	1 (100.0%)	3 (100.0%)	2 (100.0%)
Other	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

**Note:** <sup>a</sup>Four concept elicitation interview participants were identified via the online survey.  
**Abbreviations:** N, total of participants; N/A, not available.

reported were symptoms associated with hypoglycemia, including hunger (n=5, 71%), pale skin (n=5, 71%), anxiety (n=4, 57%), shakiness (n=4, 57%), and tiredness (n=4, 57%). Feeling tired, shaky, and weak were identified as the most bothersome signs/symptoms. Hypoglycemia was identified as most important to treat, followed by confusion/brain fog, low energy, and muscle development. A total of 31 impacts across six domains (physical, daily activities, work/school, sleep, emotional, and social) were reported. Common impacts were primarily cornstarch regimen-related and included restrictive diet (n=6, 86%), interrupted sleep (n=4, 57%), the need to plan ahead for diet (n=3, 43%), difficulty waking

(n=3, 43%), and limited social activities (n=3, 43%). Interrupted sleep was attributed to nighttime awakenings for cornstarch treatment, and social impacts were attributed to a regimented (“living by the clock”) and restrictive diet (cornstarch treatment limiting sugar and carbohydrate intake). Cornstarch regimen-related impacts (ie, the need to plan ahead for diet, interrupted sleep) were identified as important to treat.

### Questionnaire Construction

The draft GSD FAD was developed based on converging results from the literature review, online survey, and CE interviews (Figure 1).

The original GSDIa Morning Diary included four items that used a recall of *from the time you went to bed last night until now* and a VRS response continuum. The original GSDIa Evening Diary was comprised of 32 items, including 20 sign/symptom and 12 impact items. All items used a recall of *from the time you woke up until now* and most used a 5-point VRS response scale. Table 3 presents abbreviated item content and number of response options for each item.

**Table 3** GSD FAD Morning and Evening Diary Question Content and Summary of Revisions During Development

eDiary	Item Number	Content Covered	Response Scale	Revisions Based on Cognitive Debriefing	Item Number in Final GSD FAD
Morning	1	Number of nighttime awakenings	8-point VRS	N/A	1
	2	How refreshed felt upon waking	5-point VRS	Dropped due to overlap with item (tired after waking up)	–
	3	How difficult it was to wake up	5-point VRS	N/A	2
	4	How tired after waking up	5-point VRS	N/A	3
Evening (Symptoms)	1	Frequency of cornstarch doses during day	11-point VRS	N/A	1
	2	Feel like have low blood sugar	5-point VRS	N/A	2
	3	Tiredness	5-point VRS	N/A	3
	4	Dizziness	5-point VRS	N/A	4
	5	Shakiness	5-point VRS	N/A	5
	6	Skin feeling clammy	5-point VRS	Dropped due to misinterpretation	–
	7	Feeling hungrier than usual	5-point VRS	N/A	6
	8	Trouble concentrating	5-point VRS	N/A	7
	9	Feeling confused	5-point VRS	N/A	8
	10	Blurred vision	5-point VRS	N/A	9
	11	Sweatiness	5-point VRS	Wording revised from <i>feel sweaty</i> to <i>feel sweaty when resting or still</i>	10
	12	Feeling anxious	5-point VRS	Dropped due to misinterpretation	–

(Continued)

Table 3 (Continued).

eDiary	Item Number	Content Covered	Response Scale	Revisions Based on Cognitive Debriefing	Item Number in Final GSD FAD
	13	Feeling stressed	5-point VRS	Moved to be administered with other impact items and added attribution to GSDIa	24
	14	Forgetfulness	5-point VRS	Wording revised from <i>forgetful</i> to <i>trouble remembering things</i>	11
	15	Feeling irritable	5-point VRS	N/A	12
	16	Muscle aches	5-point VRS	N/A	13
	17	Muscle weakness	5-point VRS	N/A	14
	18	Nausea	5-point VRS	N/A	15
	19	Skin looking pale	5-point VRS	Dropped due to misinterpretation	–
	20	Trouble breathing	5-point VRS	N/A	16
Evening (Impacts)	21	Frequency of feeling different due to diet	5-point VRS	N/A	17
	22	How disruptive to plan ahead	5-point VRS	N/A	18
	23	How inconvenient to follow diet	5-point VRS	N/A	19
	24	How difficult to play sports/exercise	5-point VRS + 2 N/A options	N/A	20
	25	How difficult to perform at work/school	5-point VRS + 2 N/A options	N/A	21
	26	How bothered by appearance	5-point VRS	N/A	22
	27	Frequency of feeling annoyed	5-point VRS	Dropped due to misinterpretation	–
	28	Frequency of feeling frustrated	5-point VRS	N/A	23
	29	Frequency of feeling scared	5-point VRS	N/A	25
	30	Frequency of feeling self-conscious	5-point VRS	N/A	26
	31	Frequency of not participating in social activities	5-point VRS	N/A	27
	32	Frequency of negative effects on relationships	5-point VRS	N/A	28

**Abbreviations:** VRS, verbal rating scale; GSD FAD, Glycogen Storage Disease Functional Assessment Diary; N/A, not available.

## Combined CE and Cognitive Debriefing (CD) Interviews

Sixteen individuals (n= 9 adults aged  $\geq 18$  years, n=4 adolescents aged 12–17 years, and n=3 children aged 8–12 years) participated in the interviews (Table 4). The three children completed only the CD component due to time limits. Adult and adolescent participants most frequently reported hypoglycemia (n=5, 39%), followed by fatigue/weakness (n=4, 31%), and confusion (n=3, 23%) symptoms; and cornstarch regimen-related concerns were the most frequently reported impacts (restricted diet [n=8, 62%], planning ahead [n=5, 39%], and strict schedule [n=5, 39%]). Participants described the need for a less burdensome treatment regimen, including substantial reduction in cornstarch intake and improvements to diet. Overall, the instructions, items, and response options were interpreted as intended, and reported to be clear and relevant to most participants, with a few exceptions (Table 3).

Cognitive debriefing findings support the content validity of the GSD FAD: most participants reported no missing concepts, and the few suggested additions were infrequent, participant-specific, and did not show a consistent thematic pattern, indicating no systematic gaps in concept coverage for GSDIa.

## Questionnaire Revision

Table 3 describes item revisions and final item content following CE/CD interviews.

## Preliminary Psychometric Evaluation of the GSD FAD

The psychometric evaluation of the GSD FAD was conducted as a secondary analysis of DTX401-CL301 Phase 3 trial data.

**Table 4** Participant Characteristics for the Concept Elicitation/Cognitive Debriefing Interviews (N=16)

Characteristic	Children (n=3) <sup>a</sup>	Adolescents (n=4)	Adults (n=9)
Age (years)			
Min–Max	8–12	13–17	21–47
Mean (SD)	9.7 (2.1)	14.5 (1.7)	35.2 (10.6)
Gender (n, %)			
Female	1 (33.3%)	0 (0.0%)	6 (66.7%)
Male	2 (66.7%)	4 (100.0%)	3 (33.3%)
Ethnicity (n, %)			
Hispanic/Latino	1 (33.3%)	0 (0.0%)	0 (0.0%)
Non-Hispanic/Latino	2 (66.7%)	4 (100.0%)	9 (100.0%)
Race (n, %) <sup>b</sup>			
American Indian or Alaska Native	0 (0.0%)	0 (0.0%)	1 (11.1%)
Asian	0 (0.0%)	0 (0.0%)	0 (0.0%)
Black or African American	0 (0.0%)	0 (0.0%)	0 (0.0%)
Native Hawaiian or Other Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)
White	2 (66.7%)	4 (100.0%)	9 (100.0%)
Other	1 (33.3%)	0 (0.0%)	0 (0.0%)

**Notes:** <sup>a</sup>Participants took part in the cognitive debriefing component only. <sup>b</sup>Respondents were instructed to select all that apply.

**Abbreviation:** N, total of participants.

## Sample

Data from a total of 46 participants (n= 26 adults, n=20 adolescents and children) from global study sites were included in the psychometric evaluation (Table 5).

## GSD FAD Morning Diary and Evening Diary

The GSD FAD was administered in the DTX401-CL301 Phase 3 trial as a 3-item Morning Diary and a 28-item Evening Diary (Table 3) and was hypothesized to be comprised of three domains: (1) symptoms domain (Evening Diary Items 2–16), (2) daily impacts domain (Morning Diary Items 1–3, Evening Diary Items 1 and 17–28), and (3) sleep impacts domain (Morning Diary Items 2–3, Evening Diary Item 3). All item response scales were rescaled to range between 0 (better health) and 4 (worse health) prior to analysis.

## Assessment of Item Variability

Most items showed a high completion rate (ie, near 100%). Specifically, Morning Diary items showed completion rates ranging from 97–100% for each day of the last 14 days of screening. All Evening Diary items showed completion rates

**Table 5** Participant Characteristics for the Psychometric Evaluation (N=46)

Characteristic	Placebo (n=25)	DTX401 (n=21)	Total (N=46)
Age (years)			
Min–Max	8–37	9–32	8–37
Mean (SD)	20.8 (8.4)	20.2 (8.5)	20.5 (8.4)
Age group (n, %)			
8–17 years	10 (40.0)	10 (47.6)	20 (43.5)
≥ 18 years	15 (60.0)	11 (52.4)	26 (56.5)
Gender (n, %)			
Female	12 (48.0)	7 (33.3)	19 (41.3%)
Male	13 (52.0)	14 (66.7)	27 (58.7%)
Ethnicity (n, %)			
Hispanic/Latino	3 (12.0)	3 (14.3)	6 (13.0)
Non-Hispanic/Latino	17 (68.0)	16 (76.2)	33 (71.7)
Not reported	5 (20.0)	2 (9.5)	7 (15.2)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)
Race (n, %)			
American Indian or Alaska Native	0 (0.0)	0 (0.0)	0 (0.0)
Asian	2 (8.0)	0 (0.0)	2 (4.3)
Black or African American	0 (0.0)	0 (0.0)	0 (0.0)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)
White	23 (92.0)	18 (85.7)	41 (89.1)
Not reported	0 (0.0)	2 (9.5)	2 (4.3)
Unknown	0 (0.0)	1 (4.8)	1 (2.2)
Other	0 (0.0)	0 (0.0)	0 (0.0)

**Abbreviation:** N, total of participants.

ranging from 95–100% for each day of the last 17 days of screening, with the exception of difficult to play/exercise (range 66–83%) and difficulty at school/work (60–92%). Item response distributions showed ceiling effects on all items at each of the 14 days of screening. Average percentage of responses across Days –14 to –1 at the ceiling ranged from 29% to 38% for each item in the Morning Diary, and 0.7% to 94% for the Evening Diary. However, several items had a substantial proportion of participants (ie, 40% or more) off the ceiling, with potential to capture improvement at post-baseline visits. The most common response to items was the ceiling of each item’s rating scheme.

## Assessment of Structure and Scaling Potential

### Inter-Item Correlations

Inter-item correlations for the symptoms domain (k=15) ranged between  $0.12 \leq r_s \leq 0.72$ ; 59% of correlations met the performance criteria of  $0.30 \leq |r_s| \leq 0.90$ . Inter-item correlations for the sleep impacts domain (k=3) ranged between  $0.59 \leq r_s \leq 0.77$ ; all correlations met the performance criteria. Items correlations for the daily impacts domain (k=16) ranged between  $-0.29 \leq r_s \leq 0.83$  and 60% of correlations met the performance criteria; however, correlations between Evening Item 1 (cornstarch frequency) and other items in the domain were lower than an absolute value of 0.30.

### Corrected Item-Total Correlations

Most items showed satisfactory item discrimination (Table 6). Corrected item-total Pearson correlations ( $r_p$ ) ranged between  $0.34 \leq r_p \leq 0.76$  for the symptoms domain and between  $0.65 \leq r_p \leq 0.84$  for the sleep impacts domain. Corrected item-total Pearson correlations ( $r_p$ ) ranged between  $0.40 \leq r_p \leq 0.86$  for the daily impacts domain, except for Evening Diary Item 1 (cornstarch frequency) and Morning Diary Item 1 (nighttime awakening for cornstarch).

**Table 6** GSD FAD Item-Total Correlations and Internal Consistency Reliability: Data from Last Seven Days of Screening, Days –7 to –1

Symptoms Domain (k = 15) Cronbach's $\alpha = 0.88$	Average Corrected Item-Total Correlation	Cronbach's $\alpha$ If Item Removed
Item 8. Felt confused*	0.76	0.86
Item 7. Trouble concentrating*	0.76	0.86
Item 11. Trouble remembering*	0.74	0.86
Item 14. Felt weak*	0.72	0.86
Item 13. Muscles ached*	0.72	0.86
Item 4. Felt dizzy*	0.68	0.87
Item 2. Felt low blood sugar*	0.59	0.87
Item 3. Felt tired*	0.59	0.87
Item 12. Felt irritable*	0.53	0.87
Item 15. Felt nauseous*	0.48	0.88
Item 16. Trouble breathing*	0.40	0.88
Item 9. Blurry vision*	0.37	0.88
Item 5. Felt shaky*	0.37	0.88
Item 6. Hungrier than usual*	0.35	0.89
Item 10. Felt sweaty*	0.34	0.88

(Continued)

**Table 6** (Continued).

<b>Daily Impacts Domain (k = 14) Cronbach's <math>\alpha</math> = 0.92</b>	<b>Average Corrected Item-Total Correlation</b>	<b>Cronbach's <math>\alpha</math> If Item Removed</b>
Item 19. Difficulty following diet*	0.87	0.91
Item 18. Difficulty with planned diet*	0.87	0.91
Item 24. Felt stressed/worried*	0.86	0.91
Item 17. Felt different due to diet*	0.83	0.91
Item 23. Felt frustrated*	0.74	0.91
Item 26. Felt self-conscious/nervous*	0.67	0.91
Item 22. Bothered by appearance*	0.65	0.91
Item 28. Difficulty with relationships*	0.59	0.92
Item 25. Felt scared*	0.58	0.92
Item 21. Difficulty at school/work*	0.52	0.92
Item 27. Difficulty socializing*	0.51	0.92
Item 3. Tired after waking up**	0.51	0.92
Item 2. Difficulty waking up**	0.42	0.92
Item 20. Difficult to play/exercise*	0.41	0.92
<b>Sleep Impact Domain (k = 3) Cronbach's <math>\alpha</math> = 0.86</b>	<b>Average Corrected Item-Total Correlation</b>	<b>Cronbach's <math>\alpha</math> If Item Removed</b>
Item 3. Tired after waking up**	0.84	0.69
Item 2. Difficulty waking up**	0.73	0.80
Item 3. Felt tired*	0.65	0.87

**Notes:** \*Evening eDiary. \*\*Morning eDiary.

**Abbreviation:** GSD FAD, Glycogen Storage Disease Functional Assessment Diary.

### Internal Consistency Reliability

Each scale demonstrated a high level of internal consistency reliability, exceeding the  $\alpha \geq 0.70$  criterion (Table 6), and no item substantially increased  $\alpha$  when the item was removed from the domain. When Evening Diary Item 1 (cornstarch frequency) and Morning Diary Item 1 (nighttime awakening for cornstarch) were removed from the daily impacts domain, there was no effect on the overall Cronbach's  $\alpha$  ( $\alpha=0.92$ ). After dropping these two items, corrected item-total Pearson correlations ( $r_p$ ) for this domain improved ( $0.41 \leq r_p \leq 0.87$ ).

### Approach to Scoring

#### Stability of Daily Scores

All 7-day average scores (for Days -14 to -8 and Days -7 to -1) showed acceptable reliability; estimates exceeded ICC(A,K)  $\geq 0.70$  for both time periods examined. ICC(A,K) estimates were higher for the last seven days of screening (Days -7 to -1) (Supplementary Table 3).

Application of the Spearman-Brown Prediction formula to the ICC(A,K) estimates showed that allowing up to three days of missing data still produced 7-day average scores with adequate reliability properties ( $\geq 0.70$ ) (Supplementary Table 4).

## Scoring Algorithms

### *Symptoms Total Score (15 Items)*

A daily total score was calculated by summing Evening Diary Items 2–16 within each day, allowing for no item-level missingness (ie, if an item was missing, then the daily score was set to missing). Next, a 7-day average score was calculated, allowing up to three missing daily scores. If more than three daily scores were missing, then the 7-day average score was set to missing. The Symptoms Total Score ranges from 0 to 60, with higher scores indicating a higher frequency of symptoms.

### *Sleep Impacts Total Score (3 Items)*

A daily total score was calculated by summing Morning Diary Items 2–3 and Evening Diary Item 3 within each day, allowing for no item-level missingness. Next, a 7-day average score was generated, allowing up to three missing daily scores. If more than three daily scores were missing, then the 7-day average score was set to missing. The Sleep Impacts Total Score ranges from 0 to 12, with higher scores indicating worse sleep impacts due to GSDiA.

### *Daily Impacts Total Score (14 Items)*

A daily total score was calculated by summing Morning Diary Items 2–3 and Evening Diary Items 17–28 within each day, allowing for no item-level missingness. Next, a 7-day average score was generated, allowing up to three missing daily scores. If more than three daily scores were missing, then the 7-day average score was set to missing. The Daily Impacts Total Score ranges from 0 to 56, with higher scores indicating worse daily impacts due to GSDiA.

## GSD FAD Domain Score Descriptives

Symptoms Total Score ( $k = 15$ ) descriptives derived using Days  $-14$  to  $-8$  and Days  $-7$  to  $-1$  showed lower levels of symptom frequency and burden, respectively, with ceiling effects observed at the score-level for the Day  $-7$  to  $-1$  period.

Sleep Impacts Total Score ( $k = 3$ ) descriptives derived using Days  $-14$  to  $-8$  and Days  $-7$  to  $-1$  showed lower levels of burden, with no score-level ceiling effects for either period. Daily Impacts Total Score ( $k = 14$ ) descriptives derived using Days  $-14$  to  $-8$  and Days  $-7$  to  $-1$  showed lower levels of burden, with ceiling effects observed at the score-level in both periods (Table 7).

**Table 7** GSD FAD Domain Score Descriptive Statistics for 7-Day Periods During Screening

Score	Visit	N	Mean (SD)	Median	Q1–Q3	Min	Max	Ceiling Effect	Floor Effect
Symptoms Total Score	Baseline Minus 14 Days to Minus 8 Days	37	5.65 (4.85)	4.60	2.29–7.20	0.14	21.29	Absent	Absent
Sleep Impacts Total Score		37	3.77 (2.27)	3.75	2.14–5.60	0.00	8.43	Absent	Absent
Daily Impacts Total Score		23	8.96 (10.45)	4.86	2.13–9.78	0.60	34.14	Present	Absent
Symptoms Total Score	Baseline Minus 7 Days to Minus 1 Day	44	5.41 (5.96)	3.38	1.36–7.33	0.00	23.43	Present	Absent
Sleep Impacts Total Score		43	3.55 (2.57)	3.57	1.50–5.41	0.00	10.14	Absent	Absent
Daily Impacts Total Score		27	8.34 (10.58)	3.57	1.36–11.16	0.00	38.43	Present	Absent

**Notes:** Score level floor and ceiling effects were assessed via standardized residuals (ie, if  $|\max$  or  $\min - \text{mean}| / \text{SD} < 1$ , then severe floor or ceiling effects may be present). Ceiling effects are defined as binning of scores towards the best possible health state (ie, low GSD FAD score), and floor effects are defined as the binning of scores towards the worst possible health state (ie, high GSD FAD scores).

**Abbreviations:** GSD FAD, Glycogen Storage Disease Functional Assessment Diary; N, number of scores; SD, standard deviation; Q1, first quartile; Q3, third quartile; Min, minimum; Max, maximum.

## Reliability and Validity

### Test-Retest Reliability

The Symptoms Total Score, Daily Impacts Total Score, and Sleep Impacts Total Score all exceeded cut-offs for test-retest reliability (ICC[A,1] estimates >0.70) across a 7-day lag (Table 8).

### Convergent and Discriminant Validity

Most Spearman correlations for all GSD FAD scores showed the expected magnitude; however, the correlation between the Sleep Impacts Total Score and the Average Frequency of Nighttime Cornstarch/Glycosade Intake ( $r_s = -0.08$ ) was lower than expected (Table 9).

### Known-Groups Validity

The Symptoms Total Score distinguished between all known-groups, the Daily Impacts Total Score distinguished between one of three known-groups, and the Sleep Impacts Total Score distinguished between three of four known-groups tested (Table 10).

**Table 8** Test-Retest Reliability of GSD FAD Scores Between Two 7-Day Periods of Screening (Days -14 to -8 and Days -7 to -1) – ICC(A,1) Estimates

Score	N	Time Points	Estimate	L95	U95
Symptoms Total Score	37	2	0.85	0.73	0.92
Daily Impacts Total Score	21	2	0.96	0.91	0.98
Sleep Impacts Total Score	36	2	0.91	0.83	0.95

**Abbreviations:** GSD FAD, Glycogen Storage Disease Functional Assessment Diary; ICC, intra-class correlation coefficient; N, number of participants with complete data across time points; L95, lower bound of 95% interval for ICC(A,1); U95, upper bound of 95% interval for ICC(A,1); HO, null hypothesis tested.

**Table 9** Summary of Convergent Validity Results for GSD FAD Scores from the Last Seven Days of Screening with Co-Validating Measures

Score	Co-validator	Expected Magnitude	Result	Interpretation
Symptoms Total Score	EQ-5D-(5L/Y) Pain/Discomfort	Discriminant/Small	Medium ( $r_s = 0.39$ )	Pass
	PGIS – Had Low Blood Sugar	Medium/Large	Large ( $r_s = -0.79$ )	Pass
	PGIS – Severity of Symptom	Medium/Large	Large ( $r_s = -0.75$ )	Pass
Daily Impacts Total Score	EQ-5D-(5L/Y) Pain/Discomfort	Discriminant/Small	Medium ( $r_s = 0.33$ )	Pass
	PGIS – Bothered by Impacts	Medium/Large	Large ( $r_s = -0.81$ )	Pass
Sleep Impacts Total Score	EQ-5D-(5L/Y) Pain/Discomfort	Discriminant/Small	Large ( $r_s = 0.52$ )	Pass
	PGIS – Bothered by Impacts	Small/Medium	Large ( $r_s = -0.69$ )	Pass
	Average Frequency of Nighttime Cornstarch/Glycosade Intake	Medium	Discriminant ( $r_s = -0.08$ )	Fail
	Average Frequency of Daytime Cornstarch/Glycosade Intake	Small/Medium	Small ( $r_s = -0.26$ )	Pass

**Notes:** The GSD FAD, PGIS, starch diary (average frequency of nighttime cornstarch/Glycosade intake, average frequency of daytime cornstarch/Glycosade intake), and EuroQol-5 Dimension measures consist of both adult and pediatric versions. Analyses described herein have been performed on pooled versions of these measures. NA results were not estimable due to severe truncation of variability in the co-validator. NA results are inconclusive due to extreme ceiling effect on the co-validator.

**Abbreviations:** GSD FAD, Glycogen Storage Disease Functional Assessment Diary; PGIS, Patient Global Impression of Severity; EQ-5D-(5L/Y), EuroQol-5 Dimension (5-level/youth version);  $r_s$ , Spearman correlation; NA, not applicable.

**Table 10** Summary of Known-Groups Validity Results for the GSD FAD Scores

Score	Tests of Known-Groups									
	Frequency of Hypoglycemic Events in Past Year (< 70 mg/dl)	Frequency of Hypoglycemic Events in Past Year (< 54 mg/dl)	Average Frequency of Daytime Cornstarch/ Glycosade Intake	Average Total Daily Cornstarch/ Glycosade Intake	Number of Hours of Fasting Tolerance	EQ-5D-(5L/Y) Usual Activities	PGIS – Had Low Blood Sugar	PGIS – Overall Health	PGIS – Severity of Symptoms	PGIS – Bothered by Impacts
Symptoms Total Score	Pass	Pass	–	–	–	–	Pass	–	Pass	–
Daily Impacts Total Score	–	–	Fail	Fail	–	NA	–	–	–	Pass
Sleep Impacts Total Score	–	–	Fail	Pass	Pass	–	–	–	–	Pass

**Note:** NA result is inconclusive due to extreme ceiling effect on the co-validator.

**Abbreviations:** GSD FAD, Glycogen Storage Disease Functional Assessment Diary; PGIS, Patient Global Impression of Severity; EQ-5D-(5L/Y), EuroQol-5 Dimension (5-level/youth version).

## Minimum Detectable Change

Findings indicate that domain scores are sensitive enough to detect small score changes (Symptoms Total Score, 1.63 to 2.33 points; Sleep Impacts Total Score, 1.10 to 1.59 points; Daily Impacts Total Score, 2.06 to 2.94 points) ([Supplementary Table 5](#)).

## Discussion

The GSD FAD is a novel content-valid, disease-specific PRO for individuals ages  $\geq 8$  years with GSDIa, developed following regulatory and scientific best practices for instrument development<sup>6,12,13</sup> This research provides evidence of the GSD FAD content validity based on results from a targeted literature review, online survey, and interviews with patients; and preliminary evidence for psychometric performance based on clinical trial data.

Converging results from the targeted, concept-focused literature review, cross-sectional online survey, and CE patient interview studies identified hypoglycemia and associated symptoms of fatigue/tiredness, anxiety, hunger, shakiness, and sweating as the most frequently reported symptoms of GSDIa. Hypoglycemia was identified as most important to treat, followed by confusion/brain fog, low energy, and muscle development signs/symptoms. Results also indicate that GSDIa is highly burdensome, negatively impacting appearance, physical, emotional, sleep, and work/school function, among other domains. The most frequently reported HRQoL impacts of GSDIa were cornstarch regimen-related, including “living by the clock” to manage cornstarch and diet, a highly regimented and restrictive diet, interrupted sleep and difficulty waking due to nighttime awakenings for cornstarch treatment, and limited social activities driven by cornstarch dosing schedule and limitations for food-related activities. Individuals with GSDIa worry about experiencing hypoglycemic events despite best efforts for condition control and live in constant fear about missing a dose that may lead to serious consequences, including seizure, coma, and even death. In this research, cornstarch regimen-related impacts were identified as most important to treat.

Results of this research highlight the pervasive impacts of GSDIa on daily life, the constant daily struggle to comply with a demanding treatment regimen, and the clinical significance of assessing cornstarch treatment burden. In following, this research informed the development of a conceptual model of GSDIa that was used to draft a new PRO measure, the GSD FAD.

Qualitative CD interviews confirmed the GSD FAD content relevance, demonstrated respondent understanding of the content, and informed minor revisions to improve item clarity and eliminate redundancy.

Results from the psychometric analysis of Phase 3 DTX401-CL301 trial baseline data supported the scaling of the Symptoms Total Score, Sleep Impacts Total Score, and Daily Impacts Total Score, and yielded evidence that the domain scores are reliable and valid for the intended target population.

This research had some limitations. Variations in the patient experience of GSDIa due to culture, language, and health systems may not be fully represented in the content validity research that recruited only US participants, with study samples that included primarily White, non-Hispanic participants.

While GSD FAD items were found to be content valid for the target population, item frequency distributions identified ceiling effects for many items at baseline and the measure did not fully capture the burden that was reflected in patient experience interviews at baseline.<sup>43</sup> Ceiling effects can cause scale attenuation affecting correlations, validity estimates, score sensitivity and responsiveness, and complicate MDC. For instance, results from MDC analyses demonstrated that small changes in domain scores are interpretable; however, they are context-dependent and may be influenced by score distributions, including reduced variability due to observed ceiling effects. In this study, ceiling effects may be influenced by the clinical trial context, including selection of participants who were relatively stable and optimized for diet and blood glucose control prior to randomization, which can compress baseline score distributions and potentially underestimate responsiveness. Another possible explanation for GSD FAD ceiling effects observed in the GSDIa population studied is resilience and adaptation in the context of living with a potentially fatal disease for which there are limited treatment options. Resilience in chronic diseases has been well documented<sup>44</sup> and described as a phenomenon in which people dynamically adapt to accept their limitations, enabling them to cope, reconstruct their “internal measurement ruler”, and enhance quality of life and treatment adherence. It is possible that the GSD FAD responses

reflect the patient experience of an adapted state, and fail to capture the nuanced, contextualized patient experience that is possible to capture in interviews that use open-ended and probing questions.

In addition, the psychometric dataset was relatively small, comprised of 46 participants, which may have limited factor structure and validity analyses. This research used a CTT framework only; factor analysis was not feasible given the sample size which was generally small for a multi-domain questionnaire. Limited sample size may affect the stability of the factor structure and generalizability of the findings. Small sample size may have also affected psychometric estimates (eg, reliability and validity coefficients) which may be less stable than would be expected in larger samples. Given that some analyses involved sample sizes in the ~20–40 range, the study may be underpowered to detect small associations or subgroup differences. Also, some of the unexpected results from the convergent and known-groups analyses (eg, weaker-than-expected relationships between sleep impacts and nighttime cornstarch intake, and limited known-groups separation for daily impacts) may be due to small sample size, possibly in combination with other study limitations such as restricted variability within an optimized trial cohort, regimen tailoring, and adherence behaviors that decouple intake metrics from perceived impact, or attenuation due to ceiling effects.

Another study limitation is that this evaluation relies solely on screening and baseline data, which limits the ability to assess responsiveness, an important psychometric property requiring longitudinal data. While ceiling effects may limit responsiveness for follow-up assessment in the Phase 3 trial, this may not be a limiting factor in other samples/studies. However, this expectation requires empirical confirmation and should not be assumed without evaluating item and score distributions and responsiveness in those settings.

Results from this psychometric evaluation are preliminary since analyses were performed using Phase 3 screening/baseline data in a modest-sized and relatively selected trial cohort, which constrains conclusions related to domain structure and sensitivity to change at this stage of development. Findings may not fully generalize to broader GSDIa populations with greater variability in baseline burden, treatment practices, and disease control.

Also, administration of the GSD FAD as a daily eDiary may limit use in future research studies due to the potential response burden associated with daily completion of a multi-item measure. To address this potential limitation, an alternate version of the GSD FAD was developed that includes the same question content but applies a past 7-day recall period to each question. This alternate version of the GSD FAD may be administered on a less frequent administration schedule and can reduce response burden in trials or other research settings. Equivalence testing demonstrated comparable results between the two forms.<sup>45</sup>

Despite possible limitations, results from this preliminary psychometric evaluation were positive. Future research should evaluate the use of the GSD FAD in larger studies involving more heterogeneous populations to confirm that item relevance, response patterns, and score interpretability are consistent across different healthcare systems and treatment regimens; further examine/confirm factor structure and dimensionality; confirm that the measure is able to detect burden and scores are sensitive to changes in the clinical condition over time; and replicate and extend psychometric and interpretation analyses using longitudinal data.

To our knowledge, the GSD FAD is the first disease-specific PRO for GSDIa. Prior research has primarily used generic measures to assess the patient experience of GSDIa.<sup>5,6</sup> While generic measures are useful and capture the relative burden of GSDIa, the GSD FAD measures concepts that patients have indicated are most important and relevant to their experience of GSDIa. Disease-specific measures often yield scores with greater precision and potential to capture changes due to treatment over time than generic measures. Further, disease-specific PROs are used in various research settings to capture patient experience data. The GSD FAD can be evaluated for modification (ie, short forms, adaptive formats, age-specific versions); and as a PRO developed in alignment with regulatory guidances, has several potential applications for use, including future clinical trials involving nutritional, behavioral or educational interventions in GSDIa, natural history studies, disease monitoring programs, or clinical practice for individual patient outcomes monitoring.

## Conclusion

The development of the GSD FAD represents a significant contribution to GSDIa, a rare genetic disorder for which no disease-specific assessment currently exists. This novel content-valid PRO measures concepts that are clinically

meaningful and important from the patient perspective and yields reliable and valid scores. The next step for this assessment is to explore score responsiveness and score interpretation using longitudinal data in a larger, more heterogeneous sample.

## Abbreviations

AAV8, adeno-associated virus serotype 8; CD, cognitive debriefing; CE, concept elicitation; COA, clinical outcome assessment; CTT, Classical Test Theory; EQ-5D-(5L/Y), EuroQol-5 Dimension (5-Level/Youth Version); G6Pase, glucose 6-phosphatase; GSD FAD, Glycogen Storage Disease Functional Assessment Diary; GSDIa, Glycogen Storage Disease Type Ia; ICC, intra-class correlation coefficient; MDC, minimum detectable change; PRO, patient-reported outcome; RCI, reliable change index; SD, S=standard deviation; SEE, standard error of estimation; SEM, standard error of measurement; SEP, standard error of prediction; VRS, verbal rating scale.

## Data Sharing Statement

Deidentified participant data is available from the sponsor upon reasonable request with a signed data access agreement.

## Ethics Approval and Consent to Participate

WCG IRB (registration # IRB00000533) reviewed and approved the online survey study, concept elicitation (CE) patient interviews, and combined CE/cognitive debriefing (CE/CD) patient interviews. Prior to study participation, individuals  $\geq 18$  years provided informed consent; the parent or legal guardian of those 8 to  $<18$  years provided consent; and those 8 to  $<18$  years provided assent. Ethics approval for the DTX401-CL301 Phase 3 trial (screening/baseline data used to evaluate the GSD FAD psychometric properties) was secured through WCG IRB and at each local participating center. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

## Acknowledgments

We would like to acknowledge Maia Bowker, medical writer at Ultragenyx Pharmaceutical Inc., for her help in writing and editing this manuscript; Michelle Kelly of Ultragenyx for her review and editorial support; Ashley O'Mara of Ultragenyx for her contributions to the design of the work; and Martha Gauthier and Blaise Cureg of Lumanity for editorial support on the qualitative research. We also acknowledge the research contributions of Betsy Malkus while previously employed at Ultragenyx and Ashley Willis, Maggie Paulich and Chris Byrd while previously employed at Lumanity.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This study was funded by Ultragenyx Pharmaceutical Inc.

## Disclosure

Diane M Turner-Bowker, Jessica Butler, Deepali Mitragotri, and Alison Skrinar are employees and shareholders of Ultragenyx Pharmaceutical Inc. Shayna Egan is an employee and shareholder of Ultragenyx and was employed by Lumanity, a company paid by Ultragenyx for this work, at the time this work was conducted. Adrian Jewett and Brandon Foster are employees of Lumanity. Adrian Jewett also reports grants from Ultragenyx, during the conduct of the study. Christina Theodore-Oklota was an employee of Ultragenyx at the time this work was conducted and is currently an employee of Mahzi Therapeutics. The authors report no other conflicts of interest in this work.

## References

- Bali DS, El-Gharbawy A, Austin S, Pendyal S, Kishnani PS. Glycogen storage disease Type I. In: Adam MP, Ardinger HH, Pagon RA, et al, editors. *GeneReviews(R)*. 1993.
- Chou JY, Jun HS, Mansfield BC. Glycogen storage disease type I and G6Pase-beta deficiency: etiology and therapy. *Nat Rev Endocrinol*. 2010;6(12):676–688. doi:10.1038/nrendo.2010.189
- Kishnani PS, Austin SL, Abdenur JE, et al. Diagnosis and management of glycogen storage disease type I: a practice guideline of the American College of Medical Genetics and Genomics. *Genet Med*. 2014;16(11):e1. doi:10.1038/gim.2014.128
- Sun B, Brooks ED, Koeberl DD. Preclinical development of new therapy for glycogen storage diseases. *Curr Gene Ther*. 2015;15(4):338–347. doi:10.2174/1566523215666150630132253
- Kruger E, de Freitas HM, Ferrecchia I, Gaydon M, Lloyd A. People and families affected by Glycogen Storage Disease Type Ia (GSDIa): an analysis of narrative accounts written by individuals living with GSDIa and their caregivers. *J Health Econ Outcomes Res*. 2025;12. [Submitted]. doi:10.36469/001c.131811
- FDA. FDA patient-focused drug development guidance series for enhancing the incorporation of the patient’s voice in medical product development and regulatory decision making. Available from: <https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical>. Accessed December 30, 2024.
- Venema A, Peeks F, de Bruijn-van der Veen M, et al. A retrospective study of eating and psychosocial problems in patients with hepatic glycogen storage diseases and idiopathic ketotic hypoglycemia: towards a standard set of patient-reported outcome measures. *JIMD Rep*. 2022;63(1):29–40. doi:10.1002/jmd2.12253
- Hsu RH, Chen HA, Chien YH, et al. Bedtime extended release cornstarch improves biochemical profile and sleep quality for patients with glycogen storage disease type Ia. *Mol Genet Genomic Med*. 2023;11(10):e2221. doi:10.1002/mgg3.2221
- Storch E, Keeley M, Merlo L, Jacob M, Correia C, Weinstein D. Psychosocial functioning in youth with glycogen storage disease type I. *J Pediatr Psychol*. 2008;33(7):728–738. doi:10.1093/jpepsy/jsn017
- Derks TGJ, Rodriguez-Buritica DF, Ahmad A, et al. Glycogen storage disease Type Ia: current management options, burden and unmet needs. *Nutrients*. 2021;13(11):3828. doi:10.3390/nu13113828
- Whittal A, Merzaglia M, Nicod E. The use of patient-reported outcome measures in rare diseases and implications for health technology assessment. *Patient*. 2021;14(5):485–503. doi:10.1007/s40271-020-00493-w.
- Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health*. 2013;16(4):461–479. doi:10.1016/j.jval.2013.04.004
- EMA Committee for Medicinal Products for Human Use (CHMP). Reflection paper on the regulatory guidance for the use of the health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005:139391.
- EMA. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products; 2006. Available from: <https://www.ema.europa.eu/en/regulatory-guidance-use-health-related-quality-life-hrql-measures-evaluation-medicinal-products-scientific-guideline>. Accessed March 10, 2026.
- EMA. Reflection paper on patient experience data; 2025. Available from: <https://www.ema.europa.eu/en/patient-experience-data-ped-reflection-paper>. Accessed March 10, 2026.
- Butler J, Dress A, Theodore-Oklota C, et al. PRO71 the humanistic burden of glycogen storage disease type IA: the impact on symptoms, diet and health-related quality of life. *Value Health*. 2020;23:S341. doi:10.1016/j.jval.2020.04.1292
- Glaser BG, Strauss AL. The discovery of grounded theory. In: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine de Gruyter; 1967.
- Charmaz K. Grounded theory. In: *Rethinking Methods in Psychology*. SAGE Publications Ltd; 1995.
- Lasch KE, Marquis P, Vigneux M, et al. PRO development: rigorous qualitative research as the crucial foundation. *Qual Life Res*. 2010;19(8):1087–1096. doi:10.1007/s11136-010-9677-6
- Peterson R. *Constructing Effective Questionnaires*. SAGE Publications, Inc.; 2000.
- Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q*. 2007;71(2):287–311. doi:10.1093/poq/nfm006
- R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2022. Available from: <https://www.R-project.org>. Accessed March 28, 2026.
- Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4:1686.
- Revelle W. Package “psych” - procedures for psychological, psychometric, and personality research. Version 2.4.6. Northwestern University; 2024. Available from: <https://CRAN.R-project.org/package=psych>. Accessed March 28, 2026.
- Gamer M Lemon, J, Gamer MM et al. Package “irr” - various coefficients of interrater reliability and agreement. Version R package version 0.84.1. 2019. Available from: <https://CRAN.R-project.org/package=irr>. Accessed March 28, 2026.
- Field AP. *Discovering Statistics Using IBM SPSS Statistics*. 5th ed. Sage Publications; 2018.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281–302. doi:10.1037/h0040957
- Bland JM, Altman DG. Cronbach’s alpha. *BMJ*. 1997;314:572. doi:10.1136/bmj.314.7080.572
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334. doi:10.1007/BF02310555
- McDonald RP. The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Br J Math Stat Psychol*. 1970;23(1):1–21. doi:10.1111/j.2044-8317.1970.tb00432.x
- McGraw KO, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30. doi:10.1037/1082-989X.1.1.30
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284. doi:10.1037/1040-3590.6.4.284
- Fleiss JL. *Design and Analysis of Clinical Experiments*. Wiley Interscience; 2011.
- Griffiths P, Williams A, Brohan E. How do the number of missing daily diary days impact the psychometric properties and meaningful change thresholds arising from a weekly average summary score? *Qual Life Res*. 2022;31(12):3433–3445. doi:10.1007/s11136-022-03198-9

35. Uttl B. Measurement of individual differences: lessons from memory assessment in research and clinical practice. *Psychol Sci.* 2005;16(6):460–467. doi:10.1111/j.0956-7976.2005.01557.x
36. Cohen J. Statistical power analysis. *Curr Dir Psychol Sci.* 1992;1(3):98–101. doi:10.1111/1467-8721.ep10768783
37. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. *J Educ Stat.* 1990;15(1):84–87. doi:10.2307/1164825
38. Hattie J, Cooksey RW. Procedures for assessing the validities of tests using the “known-groups” method. *Appl Psychol Meas.* 1984;8(3):295–305. doi:10.1177/014662168400800306
39. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res.* 2005;14(2):285–295. doi:10.1007/s11136-004-0705-2
40. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes.* 2006;4:54. doi:10.1186/1477-7525-4-54
41. 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;20:1727–1736. doi:10.1007/s11136-011-9903-x
42. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59(1):12–19. doi:10.1037/0022-006X.59.1.12
43. Turner-Bowker D, Egan S, Butler J, Collis R, Mitchell J. Qualitative interviews to characterize disease and treatment burden at baseline in adult and pediatric patients participating in a pivotal phase 3 trial of DTX401 for the treatment of glycogen storage disease type Ia. Poster presented at: The Professional Society for Health Economics and Outcomes Research (ISPOR); May 13–16, 2025; Montreal, QC, Canada.
44. Sturgeon J, Zautra A. Resilience: a new paradigm for adaptation to chronic pain. *Curr Pain Headache Rep.* 2010;14(2):105–112. doi:10.1007/s11916-010-0095-9
45. Jewett A, Foster B, Mitragotri D et al. Comparison of daily vs weekly administration formats of the Glycogen Storage Disease Functional Assessment Diary (GSD FAD). Poster accepted for presentation at The Professional Society for Health Economics and Outcomes Research (ISPOR); May 17–20, 2026; Philadelphia, PA.

## Patient Preference and Adherence

**Dovepress**  
Taylor & Francis Group

### Publish your work in this journal

Patient Preference and Adherence is an international, peer-reviewed, open access journal that focusing on the growing importance of patient preference and adherence throughout the therapeutic continuum. Patient satisfaction, acceptability, quality of life, compliance, persistence and their role in developing new therapeutic modalities and compounds to optimize clinical outcomes for existing disease states are major areas of interest for the journal. This journal has been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/patient-preference-and-adherence-journal>