

# Psychometric Evaluation of the Celiac Disease Symptom Diary 2.1<sup>©</sup> Using Data from the Virtual Celiac Symptoms Study

Sonal Ghura<sup>1</sup>, Dawn Wiese Adams<sup>2</sup>, Marilyn G Geller<sup>3</sup>, Daniel A Leffler<sup>1,4</sup>, Edwin Liu<sup>5</sup>, Lori D McLeod<sup>6</sup>, Lisa M Meckley<sup>1</sup>, Muna J Tahir<sup>1</sup>, Ragy Saad<sup>1</sup>, Nicholas J Rockwood<sup>6</sup>

<sup>1</sup>Takeda Development Center Americas, Inc., Cambridge, MA, USA; <sup>2</sup>Vanderbilt University Medical Center, Nashville, TN, USA; <sup>3</sup>Celiac Disease Foundation, Woodland Hills, CA, USA; <sup>4</sup>Celiac Center, Beth Israel Deaconess Medical Center, Harvard Medical School Celiac Research Program, Boston, MA, USA; <sup>5</sup>Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO, USA; <sup>6</sup>RTI Health Solutions, Research Triangle Park, NC, USA

Correspondence: Sonal Ghura, Takeda Development Center Americas, Inc., 500 Kendall Street, Cambridge, MA, 02142, USA, Tel +608-770-9058, Email sonal.ghura@takeda.com



**Purpose:** Validated patient-reported outcome measures (PROMs) are required for use in clinical trials of celiac disease (CeD) therapies. The Celiac Disease Symptom Diary 2.1<sup>©</sup> (CSDS 2.1<sup>©</sup>), which measures the daily severity of core CeD symptoms (abdominal pain, bloating, diarrhea, nausea, tiredness), was developed according to the latest regulatory guidelines for fit-for-purpose PROMs. This study evaluated the psychometric properties of CSDS 2.1.

**Methods:** Psychometric properties of CSDS 2.1 were evaluated using data from a 12-week US observational study, the Virtual Celiac Symptoms Study (NCT05309330), in patients with CeD maintaining a gluten-free diet. Participants completed CSDS 2.1 daily and other PROMs (Patient Global Impression of Severity [PGIS], Gastrointestinal Symptom Rating Scale [GSRs], and Celiac Symptom Index [CSI]) at specified time points to evaluate the reliability, validity, and responsiveness of CSDS 2.1.

**Results:** Overall, 480 participants (338 adults, 142 adolescents) completed the study. Cronbach's alpha (baseline = 0.77 adults/adolescents) indicated high internal consistency reliability of weekly average gastrointestinal (GI; abdominal pain, bloating, nausea, diarrhea) CSDS 2.1 scores. An intraclass correlation coefficient of 0.89 (adults)/0.88 (adolescents) demonstrated high test-retest reliability among stable patients on PGIS. Moderate-to-strong correlations between weekly average GI CSDS 2.1 scores and GSRs domains at baseline and CSI at Week 3 confirmed construct validity ( $r = 0.44-0.76$ ;  $p < 0.05$ ). Weekly average GI CSDS 2.1 score changes followed expected patterns based on PGIS change groups, demonstrating responsiveness.

**Conclusion:** This evaluation provides evidence to support the use of CSDS 2.1 in clinical trials as a reliable and responsive measure of CeD symptom severity.

**Keywords:** celiac disease, celiac disease symptom diary, gastrointestinal symptoms, psychometric validation, patient-reported outcome measure, virtual celiac symptoms study

## Introduction

Celiac disease (CeD) is a systemic immune-mediated condition triggered by gluten ingestion in genetically predisposed individuals, with an estimated global prevalence of 1%.<sup>1-5</sup> CeD is characterized by inflammation of the small intestine leading to gastrointestinal (GI) symptoms such as abdominal pain, diarrhea, vomiting/nausea, and bloating/flatulence.<sup>2,6,7</sup> Individuals with CeD also commonly experience non-GI symptoms including fatigue, headache, cognitive difficulties, joint pain, and skin rash.<sup>2,7-11</sup> Further non-GI manifestations of CeD include short stature, anemia, osteopenia/osteoporosis, anxiety, and depression.<sup>5</sup>



Currently, the only management option for CeD is lifelong strict adherence to a gluten-free diet (GFD), which places a considerable burden on patients and caregivers, and many patients of varying age groups continue to experience symptoms, most often attributed to inadvertent exposure to gluten.<sup>2,7,12–15</sup>

Several treatments are under development to address the unmet needs for patients with CeD on a GFD.<sup>16</sup> To evaluate the effectiveness of these new therapies, validated patient-reported outcome measures (PROMs) are required. The majority of PROMs used in CeD clinical trials have not been developed and validated according to US Food and Drug Administration (FDA) guidelines for fit-for-purpose instruments for clinical trials and/or guidelines for CeD therapies adjunctive to a GFD.<sup>17,18</sup> As per FDA guidelines, a PROM used to construct an endpoint in CeD clinical trials should measure the core signs and symptoms of CeD with daily assessments (24-hour recall period), and have demonstrated content validity, reliability, construct validity, and an ability to detect change in relevant symptoms of CeD in the target population.<sup>18,19</sup>

The Celiac Disease Symptom Diary 2.1<sup>©</sup> (CSDS 2.1<sup>©</sup>) is a daily diary that measures the severity of typical symptoms of CeD (abdominal pain, bloating, diarrhea, nausea, and tiredness). This instrument was developed according to FDA guidelines and has demonstrated content validity in adult and adolescent populations.<sup>19–23</sup> Here we report the psychometric evaluation of CSDS 2.1, focusing specifically on validity, reliability, responsiveness, and meaningful change thresholds, using data collected in the Virtual Celiac Symptoms Study (VCSS; NCT05309330).

## Methods

### Virtual Celiac Symptoms Study (VCSS)

The VCSS (NCT05309330) was a 12-week, digital-based, observational study conducted in US adults and adolescents with CeD from July 25, 2022 to March 4, 2023. It was designed to evaluate the symptom patterns, gluten exposure, and burden of disease in patients with CeD; data on symptom severity were collected using CSDS 2.1 and other PROMs on a software application accessible via a smartphone. Patients were recruited by the Celiac Disease Foundation (CDF) via digital advertisements.

### Eligibility

Eligible patients were English speaking, were aged  $\geq 12$  years, resided in the USA, had been diagnosed with CeD for  $\geq 1$  year, had a self-reported biopsy-confirmed diagnosis of CeD (adults) or self-reported serology or biopsy-confirmed diagnosis of CeD (adolescents), were adherent to a GFD for  $\geq 6$  months, and had CeD-related symptoms (patient-reported) in the past 3 months.

### Data Collection

Data were collected prospectively during the study via a smartphone/tablet application including daily and weekly assessments ([Supplementary Table 1](#)). Patients completed the CSDS 2.1 and generic and disease-specific PROMs for comparison, including: Patient Global Impression of Severity version 1.1 (PGIS; a single item assessing GI symptom severity),<sup>24</sup> Gastrointestinal Symptom Rating Scale (GSRS),<sup>25,26</sup> and Celiac Symptom Index (CSI).<sup>27</sup> Owing to the assessment schedules of supporting measures, psychometric analyses predominantly used data from Weeks 0, 3, and 7.

### CSDS 2.1

Patients completed the CSDS 2.1 every evening for 12 weeks. CSDS 2.1 measures the severity of typical GI symptoms of CeD (abdominal pain, bloating, diarrhea, and nausea) and tiredness. In addition, three frequency items are included in a supplementary questionnaire (CSDS 2.1 - Frequency Supplement) assessing vomiting, bowel movements (all), and bowel movements classified as Type 6 or 7 on the Bristol Stool Form Scale.<sup>28</sup> Patients were asked to rate their peak severity of each symptom using a 5-point ordinal scale (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) and to count the number of each of the three frequency items during the past 24 hours.

## Individual and Composite Analyses Scores

The following scores were utilized for individual symptom analyses: average of the daily severity scores during the week for each symptom (weekly average item severity score); worst daily severity scores during the week for each symptom (weekly worst item severity score); and weekly sum of vomiting, bowel movement, and diarrhea episodes (frequency score). The following scores were utilized for composite analyses: average of the weekly scores for GI symptoms (average GI symptom severity score); average of the weekly worst item severity scores for GI symptoms (worst GI symptom severity score); average of the weekly scores for all symptoms (average total symptom severity score); and average of the worst weekly item severity scores for all symptoms (worst total symptom severity score).

## Data Analyses

All analyses, including stratification by adults ( $\geq 18$  years) and adolescents (12–<18 years), were conducted according to a prespecified statistical analyses plan. Missing data on the CDS 2.1 were reported, but not imputed. To compute the weekly average and worst severity item scores, at least 4 out of 7 days of non-missing data were required, following the conventional half-data rule for missing data in PROMs in which at least half the administration days are required to compute a score. Frequency scores were adjusted for missing data, eg for 5 non-missing days in one week, the sum was divided by 5 and multiplied by 7 to report a weekly score. To compute the weekly composite scores, weekly item scores for all items used to construct the composite scores were required. Continuous variables were summarized using measures of central tendency and categorical variables were summarized by frequency/percentage.

## Inter-Item Correlations

Inter-item correlations between the CDS 2.1 weekly average item severity, worst item severity, and frequency scores at baseline were evaluated using Pearson correlations for average severity scores, polychoric correlations for worst severity scores, Spearman correlations for frequency scores, and polyserial correlations between average and worst scores.

## Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) models were fitted to assess the structure of the weekly average and worst severity composite scores at baseline. Model fit was assessed using the comparative fit index (CFI) and non-normed fit index (Tucker–Lewis Index [TLI]).<sup>29,30</sup> For both indices, values  $>0.95$  indicated a good model fit, 0.90–0.95 marginal fit, and  $<0.90$  poor fit.<sup>31</sup> The root mean square error of approximation (RMSEA) was also evaluated; values  $<0.06$  indicated satisfactory model fit, 0.06– $<0.08$  fair fit, 0.08–0.10 mediocre fit, and  $>0.10$  poor fit.<sup>31,32</sup> For standardized root mean square residual (SRMR), values  $>0.08$  indicated poor fit.<sup>31</sup>

## Composite Score Reliability

Cronbach's coefficient alphas for the composite scores were computed at baseline. Cronbach's coefficient alphas between 0.70 and 0.90 indicate a set of items that are strongly related but not redundant.<sup>33</sup> Item-total correlations for weekly average and worst severity items were examined at baseline.

## Construct Validity (Convergent and Divergent)

Correlational analyses were conducted to examine the convergent and divergent validity of the weekly average and worst severity composite scores, and frequency items, with specified items and domains/subscales from the GSRS and the CSI at baseline and Week 3, respectively. Convergent validity refers to how strongly associated the CDS 2.1 scores are with supporting measures describing similar symptoms, while divergent validity refers to the lack of or small association between scores describing dissimilar symptoms. Moderate-to-strong positive correlations between specific CDS 2.1 items and corresponding items in the GSRS and CSI were hypothesized. Correlation coefficients ( $r$ )  $<0.30$  were considered weak, 0.30– $<0.70$  moderate, 0.70–0.90 strong, and  $>0.90$  very strong.<sup>34</sup>

## Construct Validity (Known Groups)

The ability of the CDS 2.1 to distinguish or discriminate between patient groups known to differ in disease severity was evaluated. Analysis of variance (ANOVA) models were used to examine mean differences in the symptom severity

scores (average total, worst total, average GI, and worst GI) among patients classified into subgroups (ie known groups) at baseline, based on PGIS responses: none, mild, moderate, severe, and very severe symptoms.

### Test-Retest Reliability

Test-retest reliability was used to assess the consistency of the weekly average and worst severity composite scores between baseline and Week 1 for patients who did not change on the PGIS from baseline to Week 1. A two-way mixed-effects ANOVA with absolute agreement for single measures was used to compute intraclass correlation coefficient (ICC) estimates of test-retest reliability. An ICC of 0.75 reflected good reliability.<sup>35</sup>

### Responsiveness (Ability to Detect Change)

As this was a non-interventional study, the assessment of responsiveness of CDS 2.1 was exploratory. Responsiveness was assessed by computing descriptive statistics of change in composite scores from baseline to Week 7 by the change in the PGIS score over the same period. Mean differences between baseline and Week 7 were compared using ANOVA. Effect sizes were calculated for all pairs of subgroups (eg 1+ improvement vs 1+ worsening on the PGIS) and results were reported using “collapsed” PGIS change scores in which adjacent categories with small sample sizes (ie <5) were combined.

### Thresholds of Meaningful Within-Patient Change

Anchor-based and distribution-based methods were used to establish meaningful within-patient change thresholds for the weekly CDS 2.1 average and worst severity composite scores. The anchor-based methods used a 1-point improvement from baseline in the PGIS score as the target anchor category; improvement threshold estimates were then derived using the descriptive mean or median weekly change in the CDS 2.1 severity scores from baseline to Week 7 within the target anchor category. Correlations of change that were at least 0.37 in magnitude signified that the proposed anchor measure was acceptable, based on achieving a large effect size using Cohen’s rule of thumb.<sup>36–39</sup> To support the estimates, thresholds were explored through repeated measures models using all weekly data. Each model used the change in each target item score as the dependent variable, the PGIS level as a categorical predictor, and time as a categorical covariate. Three distribution-based methods (the one-half standard deviation [SD], the standard error of measurement [SEM], and the reliable change index) were also used as supportive values.

### Software and Analysis Packages

All analyses were performed using R version 4.2 or higher within a Windows-based environment (primary analysts) or SAS version 9.4 or higher within a Linux-based environment (quality control analyst).

### Ethics Approval and Consent to Participate

The Virtual Celiac Symptoms Study was approved by the Western-Copernicus Group Institutional Review Board and was conducted in accordance with the Declaration of Helsinki. All patients (adults, parents/caregivers of adolescents, and adolescents) provided informed consent before participation.

## Results

### Patient Demographics

In total, 480 patients (338 adults, 142 adolescents) were included in the study: mean (SD) age was 31.0 (15.0) years (adults: 37.9 [12.5] years; adolescents: 14.5 [1.7] years), 82% were female, and 98% were White. The mean time since CeD diagnosis was 6.7 (4.8) years (adults: 7.3 [5.2] years; adolescents: 5.4 [3.4] years). Additional patient demographics are reported in [Supplementary Table 2](#).

### CDS 2.1 Scores

For adults and adolescents, mean item (symptom) severity scores were low across all symptoms (0.4–1.5 and 0.3–1.4, respectively) at Weeks 0, 3, and 7 ([Supplementary Table 3](#)). Mean worst item (symptom) severity scores for adults and adolescents were slightly higher at 0.9–2.3 and 0.6–2.4, respectively, across all symptoms at Weeks 0, 3, and 7

([Supplementary Table 4](#)). For both adults and adolescents, diarrhea and nausea had the lowest severity scores (mean 0.3–0.5 for diarrhea; 0.4–0.5 for nausea) and tiredness had the highest severity scores (mean 1.0–1.5) for all time points. Across Weeks 0, 3, and 7, vomiting frequency was similar for both groups ([Supplementary Table 5](#)); however, adults reported a higher weekly frequency of bowel movements than adolescents (15–16 and 12–13, respectively) and a higher frequency of diarrhea (5–7 and 2–3 for adults and adolescents, respectively; [Supplementary Table 5](#)).

Across Weeks 0, 3, and 7, average GI and average total symptom severity were similar for both adults and adolescents ([Supplementary Table 6](#)); however, worst GI and worst total symptom severity was higher for adults than adolescents ([Supplementary Table 7](#)).

## Item Response Frequencies (Overall Population)

Missing data in this study represent days in which the entire diary was not completed (no partially completed days were reported owing to data entry restrictions in the software). The mean (SD) number of days the CDSD 2.1 was not completed per patient was 0.53 (0.70), 0.99 (1.26), and 1.34 (1.64) for Weeks 0, 3, and 7, respectively. The proportions of patients with missing weekly scores in Weeks 0, 3, and 7 ranged from 3 to 10%, 13 to 16%, and 18 to 22%, respectively, across adults and adolescents. The full response scale was used by patients on all items. The proportions of patients with the lowest possible response (none) for a severity item during baseline were: diarrhea 61–74%, abdominal pain 41–46%, bloating 36–46%, nausea 65–73%, and tiredness 10–28%. The proportions of patients with the highest possible response (very severe) for a severity item during baseline were: diarrhea 0–1%, abdominal pain 1–2%, bloating 1–2%, nausea 1–2%, and tiredness 4–5%.

## CDSD 2.1 Inter-Item Correlations

Moderate-to-strong inter-item correlations ( $\geq 0.4$ ) were generally observed between the CDSD 2.1 weekly average item severity and corresponding worst item severity and frequency scores at baseline in adults ([Table 1](#)) and adolescents ([Supplementary Table 8](#)). In adults, the diarrhea severity item had the weakest correlations with the other non-diarrhea-

**Table 1** Inter-Item Correlations of CDSD 2.1 at Baseline (Adults)

CDSD 2.1 Item	Correlation Coefficient (r)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	
1. Average Diarrhea Severity	–													
2. Average Abdominal Pain Severity	0.39	–												
3. Average Bloating Severity	0.26	0.74	–											
4. Average Nausea Severity	0.23	0.56	0.50	–										
5. Average Tiredness Severity	0.24	0.54	0.53	0.45	–									
6. Vomiting Frequency	0.17	0.28	0.24	0.40	0.11	–								
7. Bowel Movements Frequency	0.58	0.21	0.15	0.17	0.12	0.18	–							
8. Diarrhea Frequency	0.72	0.33	0.23	0.22	0.13	0.24	0.65	–						
9. Worst Diarrhea Severity	<b>0.82</b>	0.45	0.29	0.25	0.24	0.18	0.47	0.64	–					
10. Worst Abdominal Pain Severity	0.34	<b>0.91</b>	0.66	0.51	0.46	0.29	0.19	0.31	0.50	–				
11. Worst Bloating Severity	0.25	0.70	<b>0.90</b>	0.50	0.47	0.28	0.12	0.21	0.35	0.74	–			
12. Worst Nausea Severity	0.21	0.53	0.46	<b>0.89</b>	0.44	0.41	0.16	0.20	0.33	0.58	0.54	–		
13. Worst Tiredness Severity	0.22	0.53	0.53	0.46	<b>0.89</b>	0.15	0.10	0.09	0.25	0.47	0.53	0.49	–	

**Notes:** Moderate-to-strong correlations were defined as  $\geq 0.4$ . Bold values indicate correlation coefficients  $> 0.80$ .

**Abbreviations:** CDSD 2.1, Celiac Disease Symptom Diary Version 2.1.

related items; correlations with abdominal pain (average, 0.39; worst 0.34) were slightly higher than correlations with other items (0.17–0.26) (Table 1). In adolescents, the correlations between the diarrhea severity item and the bloating severity (average, 0.32; worst, 0.31), nausea severity (average, 0.15; worst, 0.16), and tiredness severity items (average, 0.23; worst, 0.27) were weaker than the correlations between the diarrhea severity item and abdominal pain severity item (average, 0.38; worst, 0.43) (Supplementary Table 8).

## Confirmatory Factor Analysis (CFA)

Most of the model fit statistics for the CFAs assessing the four composite scores indicated acceptable model fit at Week 0 (CFI  $\geq$  0.95; TLI  $\geq$  0.94; SRMR  $\leq$  0.03; Table 2), but the RMSEA indicated relatively poor model fit (RMSEA  $\geq$  0.10) for the worst total symptom score during baseline. The CFA results based on the adolescent sample were generally consistent with the results from the adult sample (Supplementary Table 9).

**Table 2** Confirmatory Factor Analysis of CDS 2.1 at Baseline (Adults)

Scale/Item/Statistic	Factor Loading (SE)/Model Fit
<b>Average Total Severity</b>	
1. Average Diarrhea Severity	0.39 (0.05)
2. Average Abdominal Pain Severity	0.90 (0.02)
3. Average Bloating Severity	0.82 (0.03)
4. Average Nausea Severity	0.63 (0.04)
5. Average Tiredness Severity	0.62 (0.04)
n	338
Chi-square, <i>df</i> , <i>P</i> value	12.88, 5, 0.0246
CFI	0.99
TLI	0.97
RMSEA (90% CI)	0.07 (0.02, 0.12)
SRMR	0.02
<b>Average GI Severity</b>	
1. Average Diarrhea Severity	0.39 (0.05)
2. Average Abdominal Pain Severity	0.94 (0.03)
3. Average Bloating Severity	0.79 (0.03)
4. Average Nausea Severity	0.60 (0.04)
n	338
Chi-square, <i>df</i> , <i>P</i> value	4.87, 2, 0.0877
CFI	0.99
TLI	0.98
RMSEA (90% CI)	0.06 (0.00, 0.14)
SRMR	0.02

(Continued)

**Table 2** (Continued).

Scale/Item/Statistic	Factor Loading (SE)/Model Fit
<b>Worst Total Severity</b>	
1. Worst Diarrhea Severity	0.49 (0.05)
2. Worst Abdominal Pain Severity	0.89 (0.03)
3. Worst Bloating Severity	0.82 (0.03)
4. Worst Nausea Severity	0.68 (0.04)
5. Worst Tiredness Severity	0.60 (0.04)
n	338
Chi-square, <i>df</i> , <i>P</i> value	21.53, 5, 0.0006
CFI	0.99
TLI	0.97
RMSEA (90% CI)	0.10 (0.06, 0.14)
SRMR	0.03
<b>Worst GI Severity</b>	
1. Worst Diarrhea Severity	0.50 (0.04)
2. Worst Abdominal Pain Severity	0.94 (0.03)
3. Worst Bloating Severity	0.79 (0.03)
4. Worst Nausea Severity	0.65 (0.04)
n	338
Chi-square, <i>df</i> , <i>P</i> value	5.12, 2, 0.0771
CFI	1.00
TLI	0.99
RMSEA (90% CI)	0.07 (0.00, 0.14)
SRMR	0.01

**Abbreviations:** CDSD 2.1, Celiac Disease Symptom Diary Version 2.1; CFI, comparative fit index; CI, confidence interval; *df*, degrees of freedom; GI, gastrointestinal; RMSEA, root mean square error of approximation; SE, standard error; SRMR, standardized root mean squared residual; TLI, Tucker–Lewis Index.

## CDSD 2.1 Composite Score Reliability

At baseline, strong Cronbach's coefficient alphas ( $>0.75$ ) were observed for the four composite scores (average GI, average total, worst GI, and worst total symptom severity scores) for adults (Table 3). The strong Cronbach's coefficient alpha indicates acceptable composite score reliability. Across baseline, item-total correlations between the average GI symptom severity score and average GI severity items ranged from 0.34 to 0.77 for adults (Table 3). Similar item-total correlations were observed for the other composite scores (Table 3). Composite score reliability findings for adolescents were aligned with the data for adults (Supplementary Table 10).

**Table 3** Composite Score Reliability of CDS2 2.1 at Baseline (Adults)

Item	Average GI Severity		Average Total Severity		Worst GI Severity		Worst Total Severity	
	Item-Total, n	Alpha <sup>a</sup>	Item-Total, n	Alpha <sup>a</sup>	Item-Total, n	Alpha <sup>a</sup>	Item-Total, n	Alpha <sup>a</sup>
<b>Scale</b>		<b>0.77</b>		<b>0.80</b>		<b>0.76</b>		<b>0.79</b>
1a. Average Diarrhea Severity	0.34, 338	0.82	0.34, 338	0.83	0.51, 338	–	0.48, 338	–
2a. Average Abdominal Pain Severity	0.77, 338	0.60	0.77, 338	0.70	0.80, 338	–	0.81, 338	–
3a. Average Bloating Severity	0.66, 338	0.67	0.69, 338	0.73	0.72, 338	–	0.73, 338	–
4a. Average Nausea Severity	0.55, 338	0.73	0.57, 338	0.77	0.66, 338	–	0.67, 338	–
5a. Average Tiredness Severity	0.58, 338	–	0.58, 338	0.77	0.50, 338	–	0.63, 338	–
1b. Worst Diarrhea Severity	0.54, 338	–	0.50, 338	–	0.41, 338	0.79	0.40, 338	0.80
2b. Worst Abdominal Pain Severity	0.76, 338	–	0.73, 338	–	0.72, 338	0.63	0.72, 338	0.70
3b. Worst Bloating Severity	0.75, 338	–	0.73, 338	–	0.62, 338	0.68	0.66, 338	0.71
4b. Worst Nausea Severity	0.65, 338	–	0.64, 338	–	0.53, 338	0.73	0.56, 338	0.75
5b. Worst Tiredness Severity	0.54, 338	–	0.69, 338	–	0.51, 338	–	0.51, 338	0.76

**Notes:** <sup>a</sup>The alpha listed in bold is the alpha for the scale. The alpha listed for each item corresponds to the alpha that would be obtained if the item was removed from the scale.

**Abbreviations:** CDS2 2.1, Celiac Disease Symptom Diary Version 2.1; GI, gastrointestinal.

## CDS2 2.1 Construct Validity (Convergent and Divergent)

The majority of the estimated correlations followed the hypothesized patterns, supporting the construct validity of the composite and frequency scores. As expected, moderate-to-strong positive correlations were observed between the average GI symptom severity score and GSRS abdominal pain ( $r = 0.73$ ) and diarrhea ( $r = 0.58$ ) domains, and between worst GI symptom severity and GSRS abdominal pain ( $r = 0.74$ ) and diarrhea ( $r = 0.58$ ) domains at baseline, demonstrating convergent validity. At Week 3, moderate correlations were observed between average GI symptom severity and CSI total score ( $r = 0.68$ ) and between worst GI symptom severity and CSI total score ( $r = 0.65$ ) (Table 4). Although GI symptom severity composite scores were hypothesized to correlate weakly to moderately with the CSI total scores (to support divergent validity), the correlations tended to be more in the upper end of the moderate range (0.65–0.69) despite the recall period of the CSI (4 weeks) differing from the scoring period of the CDS2 2.1 (1 week). Nevertheless, as would be expected, these correlations were weaker than the correlations with the GSRS abdominal pain domain (0.73–0.75) giving some support to divergent validity. Similar correlations were observed among adolescents (Supplementary Table 11).

## CDS2 2.1 Construct Validity (Known Groups)

On average, higher average GI symptom severity, average total symptom severity, worst GI symptom severity, and worst total symptom severity scores were observed for subgroups of patients who reported worse states based on the PGIS at baseline for the adult population ( $p < 0.001$ ; Table 5). A similar trend was observed in the adolescent population (Supplementary Table 12).

## CDS2 2.1 Test-Retest Reliability

In the subset of patients who reported no change on the PGIS from test to retest, the ICC for average GI symptom severity was 0.89 (95% confidence interval: 0.86–0.91) overall, 0.89 (0.85–0.91) for adults, and 0.88 (0.81–0.92) for adolescents. The ICC for worst GI symptom severity was 0.77 (0.71–0.81) overall, 0.76 (0.69–0.81) for adults, and 0.78 (0.67–0.86) for adolescents. These findings indicate good test-retest reliability. Similar trends were observed for the average and worst total symptom severity scores (Supplementary Table 13).

**Table 4** Construct Validity (Convergent and Divergent) of CDSD 2.1 at Baseline and Week 3 (Adults)

Week/ Supporting Measure	Correlation Coefficient (r), Number of Patients						
	Average GI Symptom Severity	Average Total Symptom Severity	Worst GI Symptom Severity	Worst Total Symptom Severity	Vomiting Frequency	Bowel Movements Frequency	Diarrhea Frequency
<b>Baseline</b>							
GSRS Abdominal Pain	0.73*, 336	0.74*, 336	0.74*, 336	0.75*, 336	0.30*, 336	0.16*, 336	0.23*, 336
GSRS Reflux	0.50*, 336	0.50*, 336	0.43*, 336	0.45*, 336	0.20*, 336	0.23*, 336	0.23*, 336
GSRS Diarrhea	0.58*, 336	0.55*, 336	0.58*, 336	0.57*, 336	0.19*, 336	0.54*, 336	0.65*, 336
GSRS Indigestion	0.66*, 336	0.67*, 336	0.60*, 336	0.62*, 336	0.16*, 336	0.19*, 336	0.27*, 336
GSRS Constipation	0.44*, 336	0.48*, 336	0.42*, 336	0.44*, 336	0.17*, 336	-0.09, 336	-0.03, 336
<b>Week 3</b>							
CSI Overall	0.68*, 323	0.74*, 323	0.65*, 323	0.69*, 323	0.26*, 323	0.13*, 323	0.25*, 323

Note: \* $p < 0.05$ .

Abbreviations: CDSD 2.1, Celiac Disease Symptom Diary Version 2.1; CSI, Celiac Symptom Index; GI, gastrointestinal; GSRS, Gastrointestinal Symptom Rating Scale.

**Table 5** Construct Validity (Known Groups) of CDSD 2.1 at Baseline (Adults)

Known Group	Average GI Symptom Severity		Average Total Symptom Severity		Worst GI Symptom Severity		Worst Total Symptom Severity	
	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n
<b>PGIS</b>								
1 = No symptoms	0.12 (0.15)	22	0.22 (0.15)	22	0.52 (0.62)	22	0.70 (0.54)	22
2 = Mild	0.45 (0.28)	138	0.61 (0.31)	138	1.09 (0.53)	138	1.28 (0.50)	138
3 = Moderate	0.89 (0.42)	143	1.05 (0.43)	143	1.81 (0.59)	143	1.95 (0.55)	143
4 = Severe	1.55 (0.59)	30	1.68 (0.57)	30	2.74 (0.76)	30	2.83 (0.69)	30
5 = Very severe	1.68 (0.18)	4	1.97 (0.22)	4	3.50 (0.29)	4	3.55 (0.30)	4
ANOVA comparison (overall), adjusted <i>P</i> value	<0.001		<0.001		<0.001		<0.001	

Note: *P* values adjusted for multiple comparisons.

Abbreviations: ANOVA, analysis of variance; CDSD 2.1, Celiac Disease Symptom Diary Version 2.1; GI, gastrointestinal; PGIS, Patient Global Impression of Severity; SD, standard deviation.

## CDSD 2.1 Responsiveness

The mean change in CDSD 2.1 scores (average total, average GI, worst total, and worst GI) from baseline to Week 7 followed the expected pattern for all composite scores for adults whereby patients who reported greater improvement on the PGIS collapsed score (1-point or  $\geq 2$ -point change) tended to have more negative (ie improved) CDSD 2.1 change scores ( $p < 0.001$ ; Table 6). A similar trend was observed in the adolescent population for Week 7 (Supplementary Table 14).

## Thresholds of Meaningful Within-Patient Change in CDSD 2.1

A proposed within-patient improvement threshold of  $-0.31$  and  $-0.32$  points for the average GI and total symptom severity scores, respectively, was considered clinically meaningful among adults. In the context of symptom-by-day improvements, a single GI symptom improving by 1 point (eg from moderate to mild) for a single day of the week corresponds to an improvement of the average GI symptom severity score of 0.04. As such, over the course of a week,

**Table 6** CDSD 2.1 Responsiveness to Change Indicated by Change in PGIS Score (Adults)

Measure/Week	Average GI Symptom Severity		Average Total Symptom Severity		Worst GI Symptom Severity		Worst Total Symptom Severity	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
<b>PGIS, Week 7</b>								
≥2-point improvement	12	-0.63 (0.51)	12	-0.63 (0.48)	12	-1.65 (1.10)	12	-1.52 (1.02)
1-point improvement	75	-0.31 (0.38)	75	-0.32 (0.34)	75	-0.70 (0.76)	75	-0.70 (0.68)
No change	162	-0.03 (0.31)	162	-0.05 (0.31)	162	-0.20 (0.61)	162	-0.22 (0.53)
1-point worsening	56	0.19 (0.45)	56	0.13 (0.42)	56	0.25 (0.89)	56	0.17 (0.76)
≥2-point worsening	5	0.44 (0.26)	5	0.43 (0.21)	5	0.65 (0.22)	5	0.72 (0.30)
ANOVA comparison (overall), P value	<0.001		<0.001		<0.001		<0.001	

**Abbreviations:** CDSD 2.1, Celiac Disease Symptom Diary Version 2.1; GI, gastrointestinal; PGIS, Patient Global Impression of Severity; SD, standard deviation.

more than seven symptom-by-day improvements would be required to reach the 0.31 threshold for meaningful change. The proposed threshold for the worst GI and worst total symptom severity scores for adults is  $-0.75$  and  $-0.80$  points, respectively. [Supplementary Tables 15](#) and [16](#) display the mean and median change scores for the 1-point improvement groups, the least-squares mean within the repeated measures mixed model (RMMM) using a 1-point improvement on the PGIS as the anchor, as well as distribution-based estimates for each of the CDSD 2.1 composite scores. [Supplementary Figures 1](#) and [2](#) present empirical cumulative distribution function (eCDF) plots of the change from baseline to Week 7 in composite scores. eCDF curves visually demonstrate a clear distinction between average and worst change scores for those with different score changes on the PGIS, thus providing additional support for use of a 1-point change on the PGIS as an anchor for deriving meaningful within-patient change thresholds. In adolescents, for most anchor-by-score combinations, the proposed anchor was not associated with the average scores strongly enough to warrant its use as an anchor. Therefore, for all average composite scores, thresholds for meaningful within-patient change could not be determined. The meaningful within-patient change thresholds for worst GI symptom severity and worst total symptom severity scores were both estimated to be  $-0.58$  among adolescents.

## Discussion

A rigorously developed PROM such as the CDSD 2.1 has the potential to support drug developers and regulatory authorities in determining the effectiveness of new therapies to treat patients with CeD who are experiencing symptoms despite adherence to a GFD.<sup>19</sup> The aim of this study was to evaluate the psychometric properties of CDSD 2.1 using data collected in the VCSS.

In this study, the higher proportion of female versus male patients, and White versus non-White patients is consistent with the prevalence of CeD in the US population.<sup>40–42</sup> Overall, within the constraints of a non-interventional observational study in which there are minimal symptom changes throughout the duration of the study period, the average and worst CDSD 2.1 severity scores (total and GI) demonstrated strong psychometric properties among both adults and adolescents. As expected, CDSD 2.1 symptom severity scores were generally low across the study as all patients were following a GFD, the gold standard for CeD treatment. Nonetheless, participants in the VCSS utilized a range of CDSD 2.1 response options reflecting a range of CeD severity levels, and the structure of the composite scores was supported by the inter-item correlations, reliability estimates, and CFA results. High Cronbach's coefficient alphas were observed for both average and worst scores, indicating acceptable composite score reliability. The CFAs assessing the structure of the CDSD 2.1 composite scores also generally demonstrated acceptable model fit; the observed poor RMSEA model fit may have been due to the small degrees of freedom.<sup>43</sup>

Construct validity results were strong for the composite scores. Overall, correlations observed across baseline and Week 3 followed the hypothesized patterns in support of convergent validity, while directionally indicative of divergent validity of CDS 2.1. Further, higher (worse) CDS 2.1 composite scores were observed for subgroups of patients who reported worse states based on PGIS, demonstrating known-groups validity. Longitudinally, both average and worst composite scores were stable within groups of patients who were stable in their disease severity, as demonstrated through the estimates of test-retest reliability.

The mean change in composite scores across the weeks followed the expected pattern in which individuals who reported greater improvement on the PGIS tended to have a decrease (ie improvement) in CDS 2.1 scores, demonstrating the ability of CDS 2.1 to detect change. This pattern was not as clear within the adolescent sample, making it difficult to establish the responsiveness of the CDS 2.1 scores in adolescents. However, this may be a function of the smaller number of adolescent participants, the limited amount of change in disease severity across the study, the lack of therapeutic intervention, and the fact that CeD symptoms do not change spontaneously.

In addition to determining the psychometric properties of the CDS 2.1 scores in the adult and adolescent populations, thresholds of meaningful within-patient change for the composite scores were estimated. In the adult sample, anchor-based thresholds were determined for the composite scores. For the adolescent sample, anchor-based estimates could not be determined for the average severity symptom composite scores, but were determined for the worst symptom severity composite scores. However, given the difficulty with establishing adequate anchors, and the limited amount of change observed within the study, the estimated meaningful within-patient change thresholds, particularly within the adolescent sample, should be interpreted with caution. The CDS 2.1 scores should be further evaluated in the context of randomized controlled trials with a therapeutic intervention to better understand the longitudinal psychometric properties and obtain additional threshold estimates.

A notable strength of this study is the use of a mobile application for daily data collection, which enabled a large study size with minimal attrition rates/missing data, thus making the VCSS well suited to support this psychometric evaluation of the CDS 2.1. Limitations of this study include recruitment of patients via a CeD advocacy organization, reflecting a sample population who may be more knowledgeable about CeD than the wider population of patients with CeD. Additionally, biopsy-confirmed diagnosis of CeD and GFD adherence were self-reported by patients and were not confirmed by a clinician or verified through medical records. Furthermore, this study did not involve a therapeutic intervention. Hence, responsiveness to change could only be measured as a comparison between study weeks with lower or higher severity scores (presumably due to inadvertent gluten exposure) as indicated by other established PROMs. Further research is required to confirm that CDS 2.1 can robustly detect changes in CeD symptom severity in response to gluten exposure and treatment. This is currently being evaluated using data from a Phase 2 clinical study of a potential therapy for CeD (NCT05353985). Finally, although the symptoms included in CDS 2.1 are those considered most relevant in CeD, according to previous concept elicitation studies,<sup>23</sup> they do not reflect the full symptom burden of CeD; it is important to note that non-GI symptoms can also play a significant role in the health-related quality of life of patients with CeD.<sup>44,45</sup>

## Conclusions

Overall, this study provides preliminary evidence to support the use of the CDS 2.1 as a patient-reported measure of CeD symptom severity. Change in CDS 2.1 symptom severity scores has potential utility as an endpoint in clinical studies to assess the effectiveness of new therapies for CeD. Further psychometric evaluation of CDS 2.1 in a phase 2 interventional clinical study is ongoing to further assess responsiveness of the CDS 2.1 to treatment.

## Abbreviations

ANOVA, Analysis of Variance; CDF, Celiac Disease Foundation; CDS 2.1, Celiac Disease Symptom Diary 2.1<sup>®</sup>; CeD, Celiac Disease; CFA, Confirmatory Factor Analysis; CFI, Comparative Fit Index; CSI, Celiac Symptom Index; eCDF, Empirical Cumulative Distribution Function; FDA, US Food and Drug Administration; GFD, Gluten-Free Diet; GI, Gastrointestinal; GSRS, Gastrointestinal Symptom Rating Scale; ICC, Intraclass Correlation Coefficient; PGIS, Patient Global Impression of Severity version 1.1; PRO, Patient-Reported Outcome; PROM, Patient-Reported Outcome Measure;

RMMM, Repeated Measures Mixed Model; RMSEA, Root Mean Square Error of Approximation; SD, Standard Deviation; SEM, Standard Error of Measurement; SRMR, Standardized Root Mean Square Residual; VCSS, Virtual Celiac Symptoms Study.

## Data Sharing Statement

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The data sets will be provided after deidentification, in compliance with applicable privacy laws, data protection and requirements for consent and anonymization.

## Ethics Approval and Informed Consent

This study was conducted in accordance with the guidelines for Good Pharmacoepidemiology Practices by the International Society for Pharmacoepidemiology and any local regulations and approval were obtained from the institutional review board (IRB). All participant- and parent/caregiver-facing materials were reviewed and approved by the Western-Copernicus Group (WCG) IRB, including informed consent forms. To maintain confidentiality, all data collected through the online survey were de-identified.

## Acknowledgments

The authors would like to thank Lauren Nelson, Elise Matta, and Dane Korver of RTI Health Solutions for their contributions to this study. Medical writing support was provided by Ify Achebe, MSc, of Oxford PharmaGenesis, Oxford, UK and was funded by Takeda Pharmaceuticals.

The methodology and some of the results from this study have previously been presented at the 20th International Celiac Disease Symposium 2024, September 5–7, 2024, Sheffield, UK and the United European Gastroenterology Week October 12–15, 2024, Vienna, Austria.

## 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 Takeda Development Center Americas, Inc.

## Disclosure

SG, MJT, and RS are employees and shareholders of Takeda. DAL and LMM were employees and shareholders of Takeda at the time of the study. DAL is now an employee of Chugai Pharmaceutical Company. RS was employed by Jazz Pharmaceuticals, PLC until Jan 2023. DWA has served as a consultant for Takeda Pharmaceuticals and Ironwood Pharmaceuticals. EL serves as a consultant for Takeda Pharmaceuticals and is a contributing author for UpToDate celiac disease section. LDM and NJR are employees of RTI Health Solutions, which receives funding from pharmaceutical companies for patient-reported outcomes consulting services. They receive no compensation from the pharmaceutical companies and their RTI salary is not related to the projects on which they work. The authors report no other conflicts of interest in this work.

## References

1. Ludvigsson JF, Leffler DA, Bai JC, et al. The Oslo definitions for coeliac disease and related terms. *Gut*. 2013;62:43–52. doi:10.1136/gutjnl-2011-301346
2. Caio G, Volta U, Sapone A, et al. Celiac disease: a comprehensive current review. *BMC Med*. 2019;17:142. doi:10.1186/s12916-019-1380-z
3. Lindfors K, Ciacci C, Kurppa K, et al. Coeliac disease. *Nat Rev Dis Primers*. 2019;5:3. doi:10.1038/s41572-018-0054-z

4. Makharia GK, Chauhan A, Singh P, Ahuja V. Review article: epidemiology of coeliac disease. *Aliment Pharmacol Ther.* 2022;56(Suppl 1):S3–S17. doi:10.1111/apt.16787
5. Sahin Y. Celiac disease in children: a review of the literature. *World J Clin Pediatr.* 2021;10:53–71. doi:10.5409/wjcp.v10.i4.53
6. Rubio-Tapia A, Hill ID, Kelly CP, Calderwood AH, Murray JA. ACG clinical guidelines: diagnosis and management of celiac disease. *Am J Gastroenterol.* 2013;108:656–676. doi:10.1038/ajg.2013.79
7. Leffler DA, Acaster S, Gallop K, et al. A novel patient-derived conceptual model of the impact of celiac disease in adults: implications for patient-reported outcome and health-related quality-of-life instrument development. *Value Health.* 2017;20:637–643. doi:10.1016/j.jval.2016.12.016
8. Majsiak E, Choina M, Gray AM, Wysokinski M, Cukrowska B. Clinical manifestation and diagnostic process of celiac disease in Poland - comparison of pediatric and adult patients in retrospective study. *Nutrients.* 2022;14:491. doi:10.3390/nu14030491
9. Leffler DA, Green PH, Fasano A. Extraintestinal manifestations of coeliac disease. *Nat Rev Gastroenterol Hepatol.* 2015;12:561–571. doi:10.1038/nrgastro.2015.131
10. Volta U, Caio G, Stanghellini V, De Giorgio R. The changing clinical profile of celiac disease: a 15-year experience (1998–2012) in an Italian referral center. *BMC Gastroenterol.* 2014;14:194. doi:10.1186/s12876-014-0194-x
11. Poddighe D, Zhubanova G, Galiyeva D, Mussina K, Forss A. Prevalence of joint complaints in patients with celiac disease: a systematic review and meta-analysis. *J Clin Med.* 2025;14:3740. doi:10.3390/jcm14113740
12. Rubio-Tapia A, Hill ID, Semrad C, et al. American College of Gastroenterology guidelines update: diagnosis and management of celiac disease. *Am J Gastroenterol.* 2023;118:59–76. doi:10.14309/ajg.0000000000002075
13. Al-Toma A, Volta U, Auricchio R, et al. European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders. *UEG J.* 2019;7:583–613. doi:10.1177/2050640619844125
14. Payette CC, Desjardins C, Lalanne E, Marquis M, Perreault M. Exploring challenges faced by adults living with celiac disease: a food literacy perspective. *J Hum Nutr Diet.* 2025;38:e70057. doi:10.1111/jhn.70057
15. Bathrellou E, Bountziouka V, Lamprou D, et al. Higher cost of gluten-free products compared to gluten-containing equivalents is mainly attributed to staple foods. *Nutr Bull.* 2025;50:44–51. doi:10.1111/nbu.12716
16. Machado MV. New developments in celiac disease treatment. *Int J Mol Sci.* 2023;24:945. doi:10.3390/ijms24020945
17. Hindryckx P, Levesque BG, Holvoet T, et al. Disease activity indices in coeliac disease: systematic review and recommendations for clinical trials. *Gut.* 2018;67:61–69. doi:10.1136/gutjnl-2016-312762
18. US Food and Drug Administration. Celiac disease: developing drugs for adjunctive treatment to a gluten-free diet. 2022. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/ceeliac-disease-developing-drugs-adjunctive-treatment-gluten-free-diet>. Accessed January 25, 2024.
19. US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>. Accessed January 25, 2024.
20. US Food and Drug Administration. Patient-focused drug development: selecting, developing, or modifying fit-for-purpose clinical outcome assessments. 2022. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-selecting-developing-or-modifying-fit-purpose-clinical-outcome>. Accessed February 1, 2024.
21. Martin SA, Meckley LM, Harris NI, Chen KS, Leffler DA. POSA362 Symptom assessment in adolescents and adults with celiac disease. *Value Health.* 2022;25:S220. doi:10.1016/j.jval.2021.11.1073
22. US Food and Drug Administration. Patient-focused drug development: methods to identify what is important to patients. 2022. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-methods-identify-what-important-patients>. Accessed May 1, 2024.
23. Howard K, Adelman D, Ghura S, et al. Development of the celiac disease symptom diary version 2.1(©) (CDS2 2.1(©)) patient-reported outcome measure. *Qual Life Res.* 2024;33:3275–3282. doi:10.1007/s11136-024-03799-6
24. Rhatigan K, Hiron B, Kesavan H, et al. Patient global impression of severity scale in chronic cough: validation and formulation of symptom severity categories. *J Allergy Clin Immunol Pract.* 2023;11:3706–3712. doi:10.1016/j.jaip.2023.08.046
25. Revicki DA, Wood M, Wiklund I, Crawley J. Reliability and validity of the gastrointestinal symptom rating scale in patients with gastroesophageal reflux disease. *Qual Life Res.* 1998;7:75–83. doi:10.1023/A:1008841022998
26. Canestaro WJ, Edwards TC, Patrick DL. Systematic review: patient-reported outcome measures in coeliac disease for regulatory submissions. *Aliment Pharmacol Ther.* 2016;44:313–331. doi:10.1111/apt.13703
27. Leffler DA, Dennis M, Edwards George J, et al. A validated disease-specific symptom index for adults with celiac disease. *Clin Gastroenterol Hepatol.* 2009;7:1328–1334. doi:10.1016/j.cgh.2009.07.031
28. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol.* 1997;32:920–924. doi:10.3109/00365529709011203
29. Bentler PM. *EQS Structural Equations Program Manual*. Multivariate Software, Inc; 2006.
30. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika.* 2025;38:1–10. doi:10.1007/BF02291170
31. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* 1999;6:1–55. doi:10.1080/10705519909540118
32. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociol Method Res.* 1992;21:230–258. doi:10.1177/0049124192021002005
33. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press; 2015.
34. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin; 2003.
35. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155–163. doi:10.1016/j.jcm.2016.02.012
36. Hays R, Revicki D. Reliability and validity (including responsiveness). In: *Assessing Quality of Life in Clinical Trials*. New York: Oxford University Press; 2005:25–39.

37. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2:63–67. doi:10.1081/COPD-200050663
38. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–109. doi:10.1016/j.jclinepi.2007.03.012
39. Fayers PM, Hays RD. Don't middle your MID: regression to the mean shrinks estimates of minimally important differences. *Qual Life Res*. 2014;23:1–4. doi:10.1007/s11136-013-0443-4
40. Galli G, Amici G, Conti L, et al. Sex-gender differences in adult coeliac disease at diagnosis and gluten-free-diet follow-up. *Nutrients*. 2022;14:3192. doi:10.3390/nu14153192
41. Mardini HE, Westgate P, Grigorian AY. Racial differences in the prevalence of celiac disease in the US population: National Health and Nutrition Examination Survey (NHANES) 2009–2012. *Dig Dis Sci*. 2015;60:1738–1742. doi:10.1007/s10620-014-3514-7
42. Lebowhl B. Celiac disease and the forgotten 10%: the “silent minority”. *Dig Dis Sci*. 2015;60:1517–1518. doi:10.1007/s10620-015-3572-5
43. Kenny DA, Kaniskan B, McCoach DB. The performance of RMSEA in models with small degrees of freedom. *Sociol Method Res*. 2014;44:486–507. doi:10.1177/0049124114543236
44. Raiteri A, Granito A, Giamperoli A, et al. Current guidelines for the management of celiac disease: a systematic review with comparative analysis. *World J Gastroenterol*. 2022;28:154–175. doi:10.3748/wjg.v28.i1.154
45. Guennouni M, Elkhoudri N, Bourrhout A, Hilali A. Assessment of quality of life in children, adolescents, and adults with celiac disease through specific questionnaires: review. *Nutrition Clinique et Métabolisme*. 2020;34:194–200. doi:10.1016/j.nupar.2020.03.006

## Patient Related Outcome Measures

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

**Dovepress**  
Taylor & Francis Group