Original Article

Asthma Daytime Symptom Diary (ADSD) and Asthma Nighttime Symptom Diary (ANSD): Measurement Properties of Novel Patient-Reported Symptom Measures

Adam Gater, MSc[a], Linda Nelsen, MHS[b], Cheryl D. Coon, PhD[c], Sonya Eremenco, MA[d], Sean O'Quinn, MPH[e], Asif H. Khan, MBBS[f], Laurent Eckert, PhD[f], Hannah Staunton, MSc[g], Nicola Bonner, MSc[a], Rebecca Hall, MMedSci[a], Jerry A. Krishnan, MD, PhD[h], Stuart Stoloff, MD[i], Michael Schatz, MD[j], John Haughney, MD[k], and Stephen Joel Coons, PhD[d]; on behalf of the Patient-Reported Outcome Consortium's Asthma Working Group *Cheshire, Hertfordshire, and Glasgow, United Kingdom; Collegeville, Pa; St Petersburg, Fla; Tucson, Ariz; Gaithersburg, Md; Chilly-Mazarin, France; Chicago, Ill; Reno, Neb; and San Diego, Calif*

What is already known about this topic? Despite the availability of various patient-reported outcome measures, the absence of standardized measures to evaluate asthma symptom severity has been acknowledged as a barrier to clinical research and understanding patient responses to treatment.

What does this article add to our knowledge? The Asthma Daytime Symptom Diary and Asthma Nighttime Symptom Diary are new patient-reported outcome measures developed for adults and adolescents with asthma. Qualified within Food and Drug Administration's Clinical Outcome Assessment Qualification Program, evidence supporting reliability and validity of Asthma Daytime Symptom Diary and Asthma Nighttime Symptom Diary scores is presented.

How does this study impact current management guidelines? Findings from this quantitative study complement existing qualitative evidence supporting the validity of the Asthma Daytime Symptom Diary and Asthma Nighttime Symptom Diary as appropriate measures for evaluating the severity of asthma symptoms and the treatment benefit associated with therapeutic interventions.

[a]Adelphi Values Ltd, Bollington, Cheshire, United Kingdom
[b]GlaxoSmithKline, Collegeville, Pa
[c]Outcometrix, St Petersburg, Fla
[d]Patient-Reported Outcome Consortium, Critical Path Institute, Tucson, Ariz
[e]AstraZeneca, Gaithersburg, Md
[f]Sanofi, Chilly-Mazarin, France
[g]Roche Products Limited, Welwyn Garden City, Hertfordshire, United Kingdom
[h]University of Illinois at Chicago, Chicago, Ill
[i]University of Nevada, Reno, Nev
[j]Kaiser Permanente Medical Center/Kaiser Foundation Hospital, San Diego, Calif
[k]Queen Elizabeth University Hospital, Glasgow, United Kingdom

Funding for this research was provided by the following Patient-Reported Outcome Consortium members: Allergan; AstraZeneca AB, United States; Boehringer Ingelheim; Genentech; GlaxoSmithKline LLC, United States; Janssen Global Services, LLC; Merck, Sharp & Dohme Corp.; Novartis Pharma AG, United States; Pfizer, Inc; and Sanofi, France.

Critical Path Institute is supported by the Food and Drug Administration (FDA), United States of the US Department of Health and Human Services (HHS) and is 54.2% funded by FDA/HHS totaling $13,239,950, and 45.8% funded by nongovernment source(s) totaling $11,196,634. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, the FDA/HHS, or the US government. For more information, please visit FDA.gov.

Conflicts of interest: A. Gater, N. Bonner, and R. Hall report being employees of Adelphi Values Ltd who received payment from the Patient-Reported Outcome Consortium's Asthma Working Group at the Critical Path Institute to conduct the research reported in this manuscript. L. Nelsen reports being a salaried employee and stock owner of GlaxoSmithKline. C. D. Coon reports personal fees from Adelphi Values during the conduct of the study. S. Eremenco has nothing to disclose. S. O'Quinn reports being an employee of AstraZeneca during the conduct of the study. A. H. Khan reports being an employee of Sanofi during the conduct of the study. L. Eckert reports being an employee of Sanofi during the conduct of the study. H. Staunton was an employee of Adelphi Values Ltd during the conduct of this study and is now an employee and shareholder of Roche Products Limited. J. A. Krishnan reports research funding from US National Institutes of Health, Patient Centered Outcomes Research Institute, Sergey Brin Family Foundation, Regeneron, and Sanofi and is also a member of the US National Asthma Education Prevention Program 2020 Update Guideline writing committee, and the Global Initiative for Asthma. S. Stoloff has nothing to disclose. M. Schatz reports grants from Merck, ALK, and Teva and personal fees from UpToDate and the American Academy of Allergy, Asthma & Immunology outside the submitted work. J. Haughney reports personal fees from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, and Teva, outside the submitted work. S. J. Coons has nothing to disclose.

Received for publication June 23, 2021; revised October 27, 2021; accepted for publication November 17, 2021.

Available online December 8, 2021.

Corresponding author: Adam Gater, MSc, Adelphi Values Ltd, Adelphi Mill, Bollington, Cheshire SK10 5JB, UK. E-mail: adam.gater@adelphivalues.com.

2213-2198

https://doi.org/10.1016/j.jaip.2021.11.026

Abbreviations used
AASDS- Adult Asthma Symptom Daily Scales
ACT- Asthma Control Test
ADSD- Asthma Daytime Symptom Diary
ANOVA- analysis of variance
ANSD- Asthma Nighttime Symptom Diary
CFA- confirmatory factor analysis
CFI- comparative fit index
DIF- differential item functioning
FDA- Food and Drug Administration
FEV_1- forced expiratory volume in 1 second
ICC- intraclass correlation coefficient
IRT- item response theory
NNFI- non-normed fit index
NRS- numeric rating scale
PCA- principal components analysis
PEF- peak expiratory flow
PGIS- Patient Global Impression of Severity
PRO- patient-reported outcome
RMSEA- root mean square error of approximation
WRMR- weighted root mean square residual

BACKGROUND: The Asthma Daytime Symptom Diary (ADSD) and the Asthma Nighttime Symptom Diary (ANSD) were developed to meet the need for standardized patient-reported measures of asthma symptoms to assess treatment trial outcomes in adults and adolescents.
OBJECTIVE: To determine scoring and evaluate the measurement properties of the ADSD/ANSD.
METHODS: Adolescents (12-17 years) and adults (18+ years) with asthma completed draft 8-item electronic versions of the ADSD/ANSD for 10 days alongside the Adult Asthma Symptom Daily Scales (AASDS) and a Patient Global Impression of Severity (PGIS). Using classical and modern psychometric methods, initial analyses evaluated the performance of ADSD/ANSD items to inform scoring. Subsequent analyses evaluated the reliability and validity of ADSD/ANSD scores.
RESULTS: A demographically and clinically diverse sample (n = 130 adolescents; n = 89 adults) was recruited. Item performance was generally strong. However, items assessing chest pressure and mucus/phlegm demonstrated redundancy and poorer performance and were removed. Principal-components analysis, confirmatory factor analysis, and item response theory supported combining items to form 6-item total ADSD/ANSD scores. Internal consistency (α = 0.94-0.95) and test-retest reliability (intraclass correlation coefficient = 0.86-0.95) were strong. Strong correlations (r = 0.72-0.80) were observed between ADSD scores and AASDS items assessing asthma symptom frequency, bother, and impact on activities. Significant differences ($P < .001$) in mean ADSD/ANSD scores were observed between groups categorized by asthma severity (PGIS), asthma control, inhaler use, nebulizer use, activity limitations, and nighttime awakenings.
CONCLUSIONS: The ADSD/ANSD items and scores demonstrated strong reliability and validity. Implementation of the measures in interventional studies will enable the evaluation of responsiveness and meaningful within-patient change. **(J Allergy Clin Immunol Pract 2022;10:1249-59)**

Key words: *Asthma; Patient-reported outcome measure; Clinical outcome assessment; Reliability; Validity*

INTRODUCTION

Affecting more than 300 million people, asthma is the most prevalent chronic respiratory disease worldwide and among the most common chronic diseases of children and adolescents.[1,2] Asthma is characterized by recurrent episodes and exacerbations of cough, wheeze, breathlessness, and chest tightness.[3] These episodes are associated with variable airflow obstruction, often reversible spontaneously or with treatment.[3] Despite advances in the understanding of asthma and availability of disease management guidelines, the need for translating scientific progress into improved outcomes meaningful to patients with uncontrolled asthma remains high.[4,5]

Forced expiratory volume in 1 second (FEV_1) and peak expiratory flow (PEF) typically serve as standard measurements of airway function in clinical studies.[6] Eradication of or a reduction in the frequency and severity of symptoms is an indicator of asthma control and a goal of asthma management.[3,7,8] However, there is a poor correlation between objective measures of disease severity (eg, FEV_1 and PEF) and patients' self-reported experience of asthma symptoms.[9-11] Asthma symptom severity can only be evaluated by the patients themselves and, as such, well-developed patient-reported outcome (PRO) measures are necessary for characterizing these experiences in clinical studies.

In 2012, the National Institutes of Health stated that "asthma clinical research will highly benefit from standardization of major outcomes in terms of definition and assessment methodology."[12] Well-established standards exist for the development and testing of PRO measures intended for use in clinical research and practice.[13-18] However, many PRO measures historically used to assess asthma symptoms fail to meet these expectations.[19] In particular, strong evidence supporting the content validity (the extent to which the PRO measure assesses the concept of interest, ie, asthma symptom severity) and measurement properties (reliability, validity, and ability to detect change) of existing measures in adolescents and adults is generally lacking.[20,21]

To fill this gap, the PRO Consortium's Asthma Working Group at the Critical Path Institute sponsored the development of 2 new PRO measures designed to assess the severity of the core symptoms of asthma among adolescents (12 through 17 years) and adults (18 years and older). Formerly referred to collectively as the Asthma Daily Symptom Diary,[22] the Asthma Daytime Symptom Diary (ADSD) and the Asthma Nighttime Symptom Diary (ANSD) were both developed according to recommendations and scientific best practices outlined in Food and Drug Administration (FDA) guidances[13-15] and published literature.[16-18] The ADSD/ANSD were qualified by the FDA within its Clinical Outcome Assessment Qualification Program in 2019.[23] The qualification statement supports the ADSD/ANSD as measures of asthma symptoms in drug development.[24]

Results of qualitative research conducted to inform ADSD/ANSD development and provide evidence for content validity of the measures were reported previously.[22] In this study, quantitative data were collected from participants using the ADSD/ANSD to evaluate

TABLE I. Summary of analyses conducted to finalize the ADSD and ANSD conceptual frameworks, to develop preliminary scoring algorithms, and to evaluate the reliability and validity of resulting scores

Measurement property	Analysis performed	Summary of analysis
Stage 1: Item-level analyses		
Quality of completion	Frequency of missing data	Frequency of missing data for the ADSD and the ANSD (at the form and item level) over the 10-day study period was evaluated.
Item performance	Item distribution and floor and ceiling effects	The frequency and percentage of each response option of the ADSD and the ANSD was described. Floor effects and ceiling effects were assessed for each item to determine the appropriateness of ADSD and ANSD item response scales for the population and likelihood of detecting changes in symptom severity over time.
	Empirical item curves and stacked histograms	Using PGIS responses, these analyses explored whether ADSD and ANSD item response categories covered the entire range of asthma symptom severity, were ordered and distinct, and were able to distinguish between participants with different asthma severity levels.
Determining inter-item relationships	Inter-item correlations	Computed using polychoric correlation coefficients (a measure of agreement between ordinal variables)[29] defined between each pair of ADSD/ANSD items at Day 3 to ensure that items were providing distinct information. Polychoric correlation coefficients between participant responses to ADSD items (Day 3) and responses to equivalent ANSD items (Day 4) were calculated to evaluate concordance between daytime and nighttime symptoms.
Instrument structure (dimensionality)	Principal component analysis (PCA)	PCA was performed on ADSD and ANSD items at Day 3 using polychoric correlations to provide insight into the grouping of similar items together into summary scores for the ADSD and the ANSD.
	Confirmatory factor analysis (CFA)	CFA of all ADSD and ANSD items loading onto a single underlying factor was performed on the basis of data collected at Day 3 using polychoric correlations, the asymptotic covariance matrix, and robust weighted least squares estimation (ie, diagonally weighted least squares or weighted least squares with a mean and variance correction). Root mean square error of approximation < 0.06, comparative fit index > 0.95, non-normed fit index > 0.95,[30] weighted root mean square residual ≤ 1.0,[31] and factor loadings > 0.40[32] were considered indicative of good model fit.
	Item response theory (IRT)	IRT (a latent variable modeling approach) was used to gain further insight into the appropriateness of the response categories and item discrimination. Data on the ADSD and ANSD items from Days 3 and 10 (separately) were analyzed using the graded response model in FlexMIRT.[33] Item-person maps were used to evaluate the range along the underlying domain of asthma severity that the set of ADSD and ANSD items is able to reliably measure. If the ADSD and ANSD provide reliable measurement both for severe asthma and for mild or controlled asthma, the item-person map should show that the set of item thresholds produces ample information across the range of the *x*-axis. A curve showing the expected reliability of scores produced at any point along the *x*-axis was included in the plot. The curve should exceed the traditional threshold of 0.70 across most of the *x*-axis where participants are likely to be located.[34]

(continued)

TABLE I. *(Continued)*

Measurement property	Analysis performed	Summary of analysis
Stage 2: Score-level analyses		
Reliability	Internal consistency reliability	The internal consistency of the ADSD and ANSD composite scores (ie, the extent to which items belong to the same scale) was investigated using the Cronbach α coefficient based on data at Day 3 (Cronbach $\alpha > 0.70$, indicating that a set of items has enough in common to justify grouping together).[34]
	Test-retest reliability	Test-retest reliability of ADSD and ANSD scores was assessed among participants whose responses to the PGIS were the same on Day 3 and Day 10 using the intraclass correlation coefficient (ICC) as calculated using a 2-way mixed-effect model with the interaction (Shrout and Fleiss's equation 2[35] and McGraw and Wong Case 3 ICC(A,1))[36] with absolute agreement for single measures.[37]
	Differential item functioning (DIF)	Ordinal logistic regression models were used to explore DIF for the ADSD and ANSD items at Day 3 according to (1) participants' age (adolescents vs adults), (2) sex (male vs female), (3) race (White vs Black/African-American/Other), (4) ethnicity (Hispanic vs non-Hispanic), and (5) education level (some high school and below vs college or higher for adults).[38]
Construct validity	Construct validity correlations	Assessed by examining correlations of the ADSD and ANSD scores with similar concepts (ie, convergent validity) and dissimilar concepts (ie, discriminant validity) on the AASDS on Day 3.
	Known-groups method	*t*-tests and ANOVA used to evaluate the ability of ADSD and ANSD scores to differentiate between participants according to disease-related variables including asthma severity, asthma control, inhaler use, nebulizer use, nighttime awakenings, and activity limitations. *t*-tests and ANOVA were used because they have been shown to be robust to violations of normality.[39,40]
Handling of missing data	Item-level missing data	To inform the number of items within the ADSD and the ANSD that may be missing, and a reliable score still be produced, the Cronbach alpha-if-item-deleted approach was used.[41]
	Form-level missing data	To inform the number of days across which scores should be averaged to produce a reliable score, the Spearman-Brown prophecy formula was used.[42]

the cross-sectional measurement properties (eg, reliability and construct validity) of individual items and summary scores.

METHODS

Study sample

A sample of 200 participants with asthma was targeted with representation sought from 4 distinct age groups: 12 to 14 years (n = 80), 15 to 17 years (n = 40), 18 to 45 years (n = 40), and 46 years or older (n = 40). Sample size estimations sought to ensure a minimum of 10 participants per item[25] and were based on the assumption of subgroup analyses being performed in 2 groups: adolescents (12-17 years) and adults (18 years and older). An oversampling of adolescents (12-14 years) was targeted at FDA's request to evaluate item performance and the reliability and validity of ADSD/ANSD scores in this subpopulation.

Participants were recruited via referrals from primary care physicians/general practitioners or respiratory specialists from 10 geographically diverse sites throughout the United States (between August and October 2015). Eligibility criteria for this study reflect those typically used in asthma clinical trials and were consistent with criteria implemented in the qualitative research conducted during the development of the ADSD/ANSD.[22] Notably, participants were required to have a clinician-confirmed diagnosis of asthma for at least a year, to have received/filled a prescription for an asthma controller medication in the past year, and to have experienced asthma symptoms during the past 3 weeks.

To ensure recruitment of a diverse sample of participants, quotas for participant characteristics, including age, sex, ethnicity, race, education level, asthma control, recent exacerbation history (per American Thoracic Society/European Respiratory Society definitions),[26] and medication use following stepwise medication categories (as outlined by the Global Initiative for Asthma 2015 guidelines)[8] were implemented.

Study procedures

This study was conducted per the Declaration of Helsinki, and ethical approval for the study was obtained from the Copernicus Group Independent Review Board (approval code ADE2-14-418).

TABLE II. Participants' sociodemographic and clinical characteristics

Sample demographic and clinical data	12 through 14 y (n = 88)	15 through 17 y (n = 42)	18 through 45 y (n = 45)	46+ y (n = 44)	Total sample (N = 219)
Age (y)					
Mean	12.9	16.1	31.3	55	25.8
Min, Max	12, 14	15, 17	18, 45	46, 74	12, 74
Sex, n (%)					
Male	46 (52.3)	18 (42.9)	16 (35.6)	19 (43.2)	99 (45.2)
Female	42 (47.7)	24 (57.1)	29 (64.4)	25 (56.8)	120 (54.8)
Ethnicity, n (%)					
Hispanic or Latino (of any race)	33 (37.5)	16 (38.1)	13 (28.9)	14 (31.8)	76 (34.7)
Non-Hispanic or Latino	55 (62.5)	26 (61.9)	32 (71.1)	30 (68.2)	143 (65.3)
Race, n (%)					
Asian or Pacific Islander	3 (3.4)	4 (9.5)	4 (8.9)	1 (2.3)	11 (5.0)
Black/African American	26 (29.5)	12 (28.6)	14 (31.1)	13 (29.5)	65 (29.7)
Hispanic/Latino (provided as individual's race as well as ethnicity)	12 (13.6)	3 (7.1)	3 (6.7)	3 (6.8)	21 (9.6)
Other	12 (13.6)	3 (7.1)	4 (8.9)	2 (4.5)	24 (11.0)
White	34 (38.6)	20 (47.6)	22 (48.9)	22 (50.0)	98 (44.7)
Highest level of education, n (%)					
Some high school	—	—	11 (24.4)	19 (43.2)	30 (33.7)
College or higher	—	—	33 (73.3)	25 (56.8)	58 (65.2)
Missing data	—	—	1 (2.2)	—	1 (1.1)
ACT, n (%)					
Well-controlled (≥20)	33 (37.5)	14 (33.3)	15 (33.3)	18 (40.9)	80 (36.5)
Not well-controlled (16-19)	24 (27.3)	16 (38.1)	14 (31.1)	11 (25.0)	65 (29.7)
Very poorly controlled (≤15)	31 (35.2)	12 (28.6)	16 (35.6)	15 (34.1)	74 (33.8)
Experience of an exacerbation in the 2 wk before screening according to physician,[26] n (%)					
No exacerbation	59 (67.0)	23 (54.6)	34 (75.6)	26 (59.1)	142 (64.8)
Moderate exacerbation	26 (29.6)	16 (38.1)	8 (17.8)	16 (34.1)	66 (30.1)
Severe exacerbation	3 (3.4)	3 (7.1)	2 (4.4)	3 (6.8)	11 (5.0)
Medication step as prescribed by physician,[8] n (%)					
Step 1	32 (36.4)	10 (23.8)	10 (22.2)	8 (18.2)	60 (27.4)
Step 2	18 (20.5)	10 (23.8)	10 (22.2)	9 (20.5)	47 (21.5)
Step 3	17 (19.3)	9 (21.4)	9 (20.0)	9 (20.5)	44 (20.1)
Step 4	16 (18.2)	9 (21.4)	11 (24.4)	14 (31.8)	50 (22.8)
Step 5	5 (5.7)	4 (9.5)	5 (11.1)	4 (9.1)	18 (8.2)

Informed consent (parental/guardian permission and assent for participants 12 -17 years) was obtained for all participants before entry into the study.

All eligible participants attended an initial visit with site staff (within 5 working days of screening) during which they received training on how to operate a handheld device (HTC H2 smartphone) on which the study outcome measures were completed and to explain study requirements. A short interview was conducted with each participant to ensure they understood the training.

During the study, participants were expected to complete the 8 ADSD items every evening and the 8 ANSD items every morning for 10 days. Alongside the ADSD/ANSD, participants completed items assessing limitations in physical activities (evening only), nighttime awakenings (morning only), and rescue inhaler/nebulizer use (morning and evening). Additional PRO measures completed during the 10-day study period included a Patient Global Impression of Severity (PGIS) item and the Adult Asthma Symptom Daily Scales (AASDS).[27] Alerts and alarms were programmed into the devices to remind participants to complete assessments at the relevant time points. Automated alerts informed study investigators of consecutive missed assessments (to follow-up with participants). Participants missing 2 consecutive assessment days (2 full days or 4 consecutive assessments) were excluded from the analyses. Skipping of individual items was allowed, but participants were asked to confirm they wished to proceed without providing an answer as a means of monitoring intentional skips.

On completion of the 10-day study period, participants attended a second study site visit to return the study device.

Study outcome measures

The ADSD and ANSD are PRO measures designed to assess self-reported severity of the core, defining symptoms of asthma. Draft ADSD/ANSD versions completed during this study had 8 items each assessing the following symptoms: difficulty breathing, wheezing, shortness of breath, chest tightness, chest pressure, chest pain, cough, and mucus (phlegm).[22] The ADSD is designed to be completed before going to bed and asks respondents to rate the severity of the asthma symptoms during the day. Conversely, the

TABLE III. CFA factor loadings and fit statistics for the ADSD and the ANSD (at Day 3)

	ADSD (n = 196)				ANSD (n = 197)			
	Original (8-item)		Final (6-item)		Original (8-item)		Final (6-item)	
Item	Unidimensional factor	Fit statistics	Unidimensional factor	Fit statistics	Unidimensional factor	Fit statistics	Unidimensional factor	Fit statistics
1. Difficulty breathing	0.924	RMSEA = 0.226	0.941	RMSEA = 0.144	0.945	RMSEA = 0.165	0.954	RMSEA = 0.135
2. Wheezing	0.855		0.880		0.924		0.932	
3. Shortness of breath	0.912	**CFI = 0.984**	0.930	**CFI = 0.994**	0.935	**CFI = 0.992**	0.947	**CFI = 0.996**
4. Pressure (heavy feeling) on chest	0.967				0.963			
5. Chest tightness	0.928	**NNFI = 0.978**	0.857	**NNFI = 0.990**	0.950	**NNFI = 0.989**	0.896	**NNFI = 0.994**
6. Chest pain	0.855		0.861		0.898		0.903	
7. Cough	0.861	WRMR = 1.303	0.828	**WRMR = 0.655**	0.826	**WRMR = 0.982**	0.818	**WRMR = 0.618**
8. Mucus/phlegm	0.828				0.779			

CFI, Comparative fit index; *RMSEA*, root mean square error of approximation; *NNFI*, non-normed fit index; *WRMR*, weighted root mean square residual.
Bolded values indicate fit statistics within accepted thresholds for good model fit.

ANSD is designed to be completed on waking and asks respondents to rate the severity of the asthma symptoms during the night. The ADSD and ANSD ask respondents to rate the severity of their asthma symptoms at their worst using a 0 to 10 numeric rating scale (NRS) ranging from "None" to "As bad as you can imagine."

At the initial site visit, all participants completed the Asthma Control Test (ACT) on paper to provide an independent measure of asthma control. For this study, levels of asthma control were characterized according to empirically confirmed cutoffs: well-controlled asthma (ACT score, ≥20); not well-controlled asthma (ACT score, 16-19); and very poorly controlled asthma (ACT score, ≤15).[28]

The PGIS was completed by participants every day using the study device (once in the morning and once in the evening) throughout the 10-day study period. The PGIS is a single item that asks respondents to rate their asthma symptoms ("Overall, please rate your asthma symptoms since you went to bed last night/since you got up this morning.") using a 0 to 10 NRS (0 = "No symptoms" to 10 = "As bad as you can imagine").

The AASDS was completed by participants on Day 3 and Day 10 on the study device. The AASDS was selected because it has been previously implemented in clinical studies and includes both a daytime symptom scale with 4 items (frequency of asthma symptoms, bother associated with asthma symptoms, levels of activity, and impact of symptoms on activity) and a nocturnal scale (composed of a single item assessing nighttime awakening due to asthma symptoms).[27] All daytime symptom items are assessed using a 0 to 6 NRS.

Statistical analyses

Analyses were conducted in 2 distinct stages. During Stage 1, item-level analyses were conducted to assess the performance of individual ADSD/ANSD items and the relationships between items to finalize measure content and develop preliminary scoring algorithms. In Stage 2, score-level analyses were conducted to evaluate the reliability and validity of ADSD/ANSD scores.

Following Stage 1, a meeting was convened to make decisions regarding ADSD/ANSD item content (including the deletion of poorly performing or redundant items) and scores to be evaluated during Stage 2. This meeting was held on January 28, 2016, and attended in-person by representatives from the Critical Path Institute, the clinical expert panel, sponsor firms, and the instrument development team; FDA representatives joined via teleconference.

The analyses used to evaluate item performance and cross-sectional measurement properties of the ADSD/ANSD were conducted using classical and modern psychometric methods. An overview of analyses involved in each stage is provided in Table I. Unless specified otherwise, all analyses were conducted on the cross-sectional analysis sample (ie, participants with available data at Day 3). Day 3 was selected to allow participants the first 2 days to familiarize themselves with the ePRO device and study measures, also providing an opportunity for any issues with completion (should they arise) to be resolved. No imputations were performed for item-level missing data. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC),[43] Mplus (Muthen & Muthen, Los Angeles, CA),[44] and FlexMIRT (Vector Psychometric Group, Chapel Hill, NC).[33]

RESULTS

Participants' sociodemographic and clinical characteristics

A total of 219 adolescents (n = 130) and adults (n = 89) with mild to severe persistent asthma were enrolled. All prespecified recruitment quotas (excluding step 5 medications) were met, as

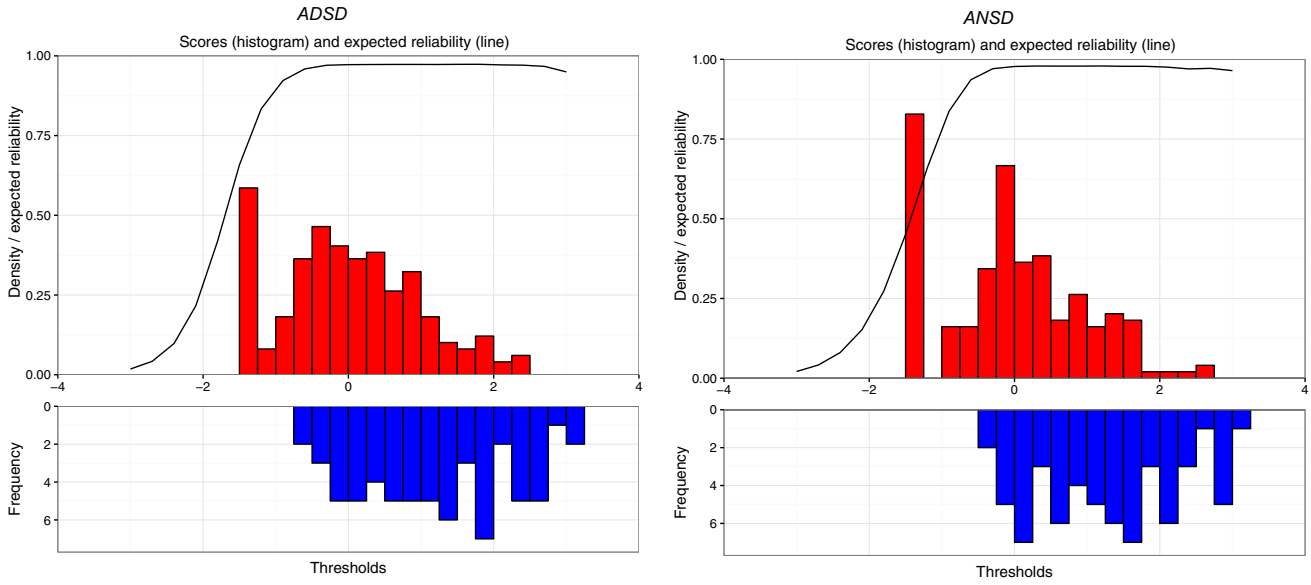

FIGURE 1. Item response theory item-person maps for ADSD and ANSD at Day 3 (n = 196).

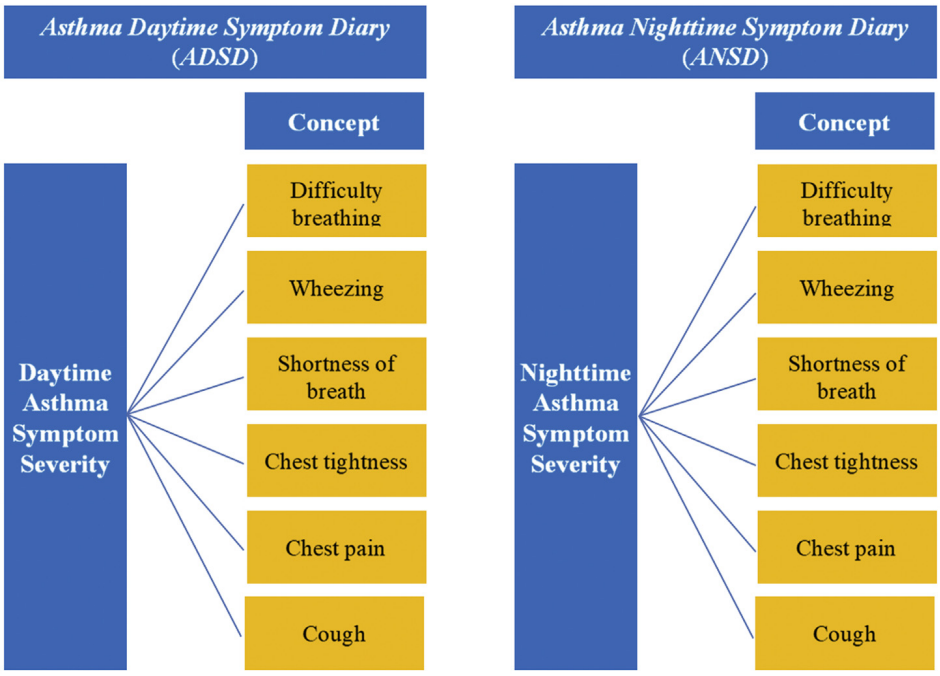

FIGURE 2. ADSD and ANSD conceptual frameworks.

reflected in the sample's diverse sociodemographic and clinical characteristics (Table II).

Stage 1 (item-level analyses)

Quality of completion. Levels of completion for the ADSD/ANSD were high, with missed completions at the form level of less than 10% of the sample for each day. The frequency of missing data at the item level was also very low, with more than 98% of participants completing all items of the ADSD/ANSD at each scheduled assessment. No evident patterns were observed among those participants who missed completions or skipped items.

Item performance. Response frequency distributions for each ADSD/ANSD item across Days 3 to 10 highlighted a substantial proportion of participants (≥35%) scoring at the ceiling (ie, selecting 0 or "None"). Ceiling effects were generally higher for the ANSD (reflecting on nighttime symptoms) than for the ADSD (reflecting on daytime symptoms). However, despite the presence of ceiling effects, observation of participant responses to items highlighted endorsement of response

TABLE IV. Construct validity: Pearson correlations between the 6-item ADSD and ANSD with AASDS items on Day 3 (n = 212)

AASDS item	ADSD	ANSD
How often did you experience symptoms today?	0.721	0.538
How much did your asthma symptoms bother you today?	0.802	0.489
How much activity could you do today?	0.479	0.231
How often did your asthma affect your activities today?	0.750	0.518
Did you wake up with asthma symptoms?	0.375	0.528

categories across the NRS continuum. The proportion of participants selecting each response option among adult and adolescent subgroups was similar.

Empirical item curves and stacked histograms indicated that individual response categories for each item of the ADSD/ANSD were generally distinct and responses to the items were generally consistent with reports of symptom severity (based on PGIS).

Inter-item relationships. Polychoric correlation coefficients between pairs of items within the ADSD/ANSD at Day 3 revealed a well-related set of items, with all correlations exceeding 0.6. Correlations between item 4 ("Pressure (heavy feeling) on your chest") and item 5 ("Chest tightness") were very high ($r \geq 0.93$), suggesting redundancy. Correlation coefficients between items on the ANSD (15 of 28 inter-item correlations exceeding 0.8) were generally stronger than for the ADSD (9 of 28 inter-item correlations exceeding 0.8). Moderate to strong ($r = 0.6$-0.8) correlations were observed between responses to the ADSD on Day 3 and the ANSD on Day 4, indicating concordance between daytime and nighttime symptoms.

Instrument structure. Principal-component analysis (PCA) for the ADSD/ANSD indicated a single dimension as best capturing the variability in responses at Day 3 (dominant eigenvalues of 6.42 and 6.62 compared with the next largest values of 0.52 and 0.48 for the ADSD and ANSD, respectively). Confirmatory factor analysis (CFA) conducted separately for the ADSD/ANSD indicated that all items loaded strongly onto the single factor (>0.70). Factor loadings for each measure were strongest for item 4 ("Chest pressure") and weakest for item 8 ("Mucus/phlegm") (Table III). Some model fit statistics indicated good model fit for the ADSD/ANSD, although some model fit statistics did not meet prespecified thresholds and suggested room for model improvement. Modification indices for 2-item combinations, item 4 ("Chest pressure") with item 5 ("Chest tightness") and item 7 ("Cough") with item 8 ("Mucus/phlegm"), exceeded the prespecified threshold, which may indicate item redundancy.

The unidimensional graded response item response theory (IRT) models also indicated good marginal fit for each item on the ADSD/ANSD, suggesting that each item is a good measure of asthma symptoms and the response categories are appropriately spaced. Similar to the CFA, some potential item redundancy was identified via several pairs of items exceeding prespecified thresholds for local dependence. Item-person maps developed as part of the IRT analysis indicated that the reliability of scores among asymptomatic participants was low (with this being more pronounced for the ANSD), making it difficult to distinguish between participants with no symptoms and those with mild symptoms (which may have important implications for deriving certain endpoints, eg, "symptom-free days"). Reliability among moderate to severe participants was high (Figure 1).

Item finalization and development of scoring algorithms

Following review of evidence from the item-level (Stage 1) analyses, the previous qualitative research, and discussion with FDA, the consensus decision was to remove item 4 ("Chest pressure") and item 8 ("Mucus/phlegm") from the ADSD/ANSD. Reasons for the removal included item redundancy (item 4 with item 5 assessing chest tightness and item 8 with item 7 assessing cough); poor item performance (eg, CFA factor loadings and modification indices); sparse evidence of importance of concepts to patients (ie, few participants spontaneously mentioning concepts and reporting concepts to be relevant during qualitative interviews); and inadequate clinical relevance. Furthermore, chest pressure is considered more indicative of cardiac issues, and mucus/phlegm is not regarded as a "core" symptom of asthma.

Based on observed differences between ratings of daytime and nighttime asthma symptoms (nighttime symptoms typically of lesser severity than daytime symptoms), there was consensus that separate scores should be calculated for the ADSD and the ANSD. In both cases, it was agreed that an average score across all 6 remaining items ranging from 0 to 10 would be calculated, with each symptom contributing equally to the measurement of overall asthma symptom severity (Figure 2).

For this study, summary scores were derived and tested using an average of ADSD/ANSD scores across multiple days (ie, Days 3 to 10) to reflect weekly asthma symptom severity. The Cronbach alpha-if-item-deleted approach highlighted that even when the 2 least reliable items of the ADSD (item 2 wheezing and item 5 chest tightness) and ANSD (item 5 chest tightness and item 7 cough) were evaluated, internal consistency reliability was acceptable ($\alpha > 0.7$). Internal consistency reliability was strongest when 3 or more items were considered ($\alpha \geq .85$). Scores from a single day were also found to provide reliable approximations of mean ADSD/ANSD scores from Days 3 through 10 (intraclass correlation coefficient > 0.7). However, reliability increased markedly when data from 2 days were considered (intraclass correlation coefficient > 0.8) and was even stronger when 4 or more days were considered (intraclass correlation coefficient > 0.9). Based on these reliability estimates, a greater than 50% rule was used in this study, with scores calculated only if data for 4 or more items were available in a day and, subsequently, if 4 or more days data were available across Days 3 to 10.

Stage 2 (score-level analyses)

Instrument structure. PCA for the revised 6-item ADSD/ANSD again indicated a single dimension as best capturing the variability in responses at Day 3 (dominant eigenvalues of 4.88 and 5.12 compared with the next largest values of 0.37 and 0.32 for the ADSD and ANSD, respectively). Component CFA factor loadings indicated that all 6 items of both the ADSD and the ANSD loaded strongly (>0.8) onto the single factor. Fit statistics for the ADSD/ANSD were generally supportive of good model fit and improved compared with the original 8-item versions of both (Table III).

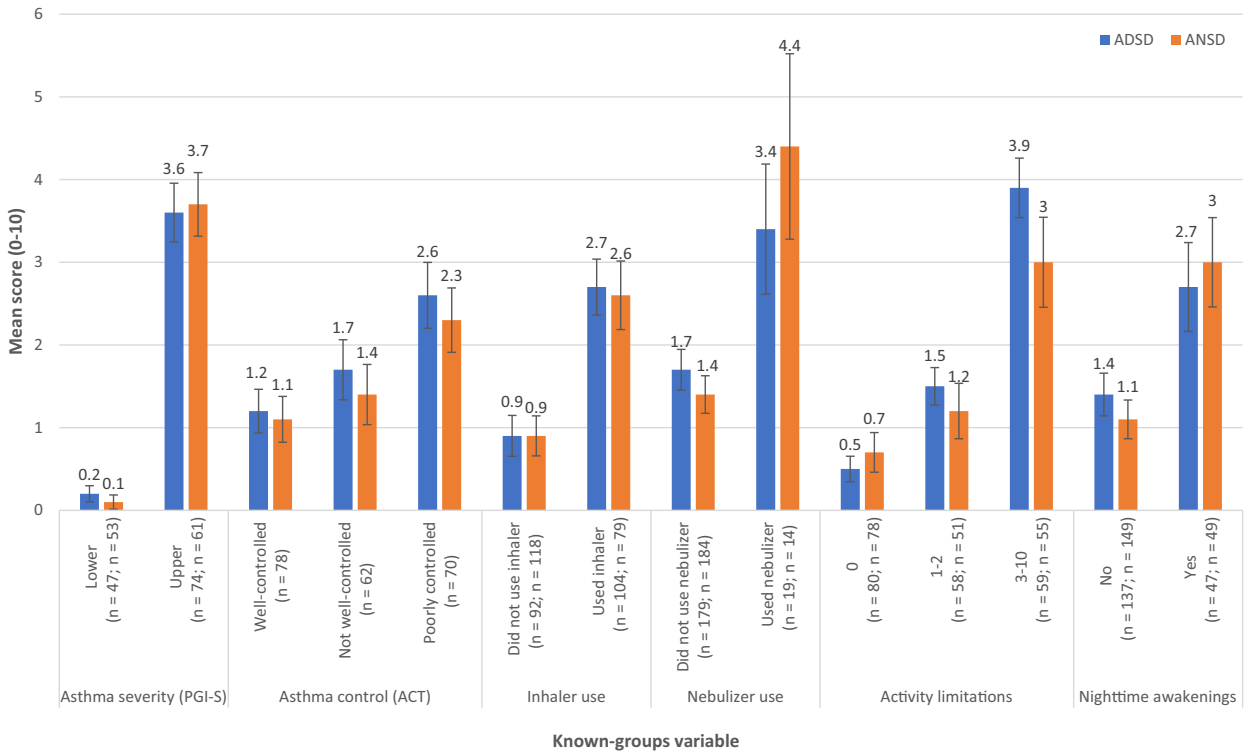

FIGURE 3. Summary of construct validity of ADSD and ANSD scores using the known-groups method ($P < .05$ for all comparisons within variable categories).

IRT marginal item fit statistics for the ADSD/ANSD were again supportive of model fit. Some item pairs still exceeded pre-specified thresholds for local dependence. Although this suggests item redundancy, these elevated fit statistics generally corresponded to pairs of breathing symptoms (eg, difficulty breathing and wheezing) or chest symptoms (eg, chest pain and chest tightness), where some relationship would be expected.

Reliability. Internal consistency of scores for the 6-item ADSD/ANSD was high (Cronbach $\alpha = 0.94$ and 0.95, respectively). In addition, scores for the 6-item ADSD/ANSD demonstrated "good" test-retest reliability (intraclass correlation coefficient = 0.86 [95% CI, 0.78-0.91] and 0.95 [95% CI, 0.92-0.97], respectively) based on participants defined as stable between Day 3 and Day 10 according to PGIS ratings (ie, participants who selected the same severity rating at each time point).

Differential item functioning. Differential item functioning (DIF) was not observed for any of the ADSD items according to the explored parameters. DIF was also not observed for any ANSD items according to participant age, sex, ethnicity, or education. Uniform DIF was evident for item 6 ("Chest pain") on the ANSD when participants were categorized according to race. Because the total number of DIF analyses elevates the probability of Type 1 error occurring by chance alone, this small amount of observed DIF is considered acceptable.

Construct validity. Table IV presents correlations between ADSD/ANSD items and AASDS items. Strong correlations were observed, as expected, between the ADSD and the AASDS items relating to frequency ($r = 0.72$), asthma symptom bother ($r = 0.80$), and impact on activities ($r = 0.75$). Correlations between the ANSD and AASDS items assessing similar concepts were slightly weaker ($r = 0.49$-0.54), which may be explained by differences in the recall periods implemented by the measures ("since you went to bed last night" and "today," respectively).

Statistically significant differences ($P < .001$) in mean ADSD/ANSD scores were observed between groups categorized according to self-reported asthma severity (PGIS), asthma control, inhaler use, nebulizer use, activity limitations, and nighttime awakenings (Figure 3). This indicates that the ADSD/ANSD can distinguish between groups that are known to differ.

DISCUSSION

Patient-reported evaluations of asthma symptom severity are important outcomes in asthma treatment research. Study sponsors have used various PRO measures to evaluate symptom severity. However, the limitations of PRO measures historically implemented in clinical trials have precluded the selection of a standardized measure that can be used by study sponsors across clinical programs, limiting the interpretation of symptom-based endpoints and the ability to compare data across studies.[19] The ADSD/ANSD were developed as part of a multi-sponsor collaboration, alongside clinical experts and in discussion with FDA to fill that measurement gap.

Findings from this quantitative study complement existing qualitative evidence supporting the content validity of the ADSD/ANSD.[22,45] Strong support was found for the reliability and validity of ADSD/ANSD items and scores derived from both measures. In seeking to develop standard measures for use across

clinical trials in patients with a clinical diagnosis of mild persistent through severe persistent asthma, care has been taken to ensure involvement and testing of the ADSD/ANSD in sociodemographically (eg, age, sex, race, ethnicity, and education level) and clinically (eg, asthma severity and medication use) diverse samples.

Establishing cross-sectional measurement properties is an important milestone for newly developed PRO measures to provide confidence that measures perform as intended. The ADSD/ANSD are being administered over time in multinational clinical studies to generate additional evidence of their longitudinal measurement properties (including further evidence of quality of completion, compliance, item response/score distributions, reliability, and construct validity). Data collected in these interventional settings will also be used to evaluate the ability of ADSD/ANSD scores to detect changes in symptom severity and to determine the amount of within-patient change that is clinically meaningful. Furthermore, the implementation of "washout periods" in clinical studies provides the opportunity to explore measurement properties among participants generally expected to be more severe (ie, greater proportion on step 5 medications) and to have greater variability in asthma severity than those typically recruited into a non-interventional study.

As the ADSD/ANSD are used in clinical studies it will be necessary to consider how they may be used to derive efficacy endpoints. The ADSD/ANSD could be used to derive symptom-free days, a concept commonly used in clinical trials but for which there are concerns regarding lack of standardization and of sensitivity to change (due to floor/ceiling effects) in participants with severe and mild asthma.[19] However, by providing a more comprehensive and detailed means of assessing asthma symptom severity, they could also be used to derive alternative endpoints potentially of greater relevance and more meaningful to patients (eg, meaningful reduction in symptom severity). Ultimately, the endpoints derived from the ADSD/ANSD may differ depending on the context in which the measures are to be used (eg, population, trial design, and investigational product).

CONCLUSIONS

Strong evidence supports the reliability and validity of ADSD/ANSD items and scores. Endorsed by FDA as measures of asthma symptom severity in drug development, the ADSD/ANSD are currently being implemented in clinical studies. Future research will seek to generate evidence regarding the responsiveness and meaningful within-patient score change, which will be important for establishing the use of the ADSD/ANSD as standardized efficacy endpoint measures in confirmatory studies.

Acknowledgments

We thank the members of the US Food and Drug Administration's Qualification Review Team for its feedback during the development of the ADSD and the ANSD.

Further information about the ADSD and the ANSD, and how to access these measures, is available at http://www.c-pathcoas.org/.

REFERENCES

1. Lai CK, Beasley R, Crane J, Foliaki S, Shah J, Weiland S, et al. Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). Thorax 2009;64:476-83.
2. The Global Asthma Report 2018. Auckland, New Zealand: Global Asthma Network. Accessed July 29, 2020. http://www.globalasthmareport.org/Global%20Asthma%20Report%202018.pdf
3. National Asthma Education and Prevention Program. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma—Summary Report 2007. J Allergy Clin Immunol 2007;120:S94-138.
4. Stanford RH, Gilsenan AW, Ziemiecki R, Zhou X, Lincourt WR, Ortega H. Predictors of uncontrolled asthma in adult and pediatric patients: analysis of the Asthma Control Characteristics and Prevalence Survey Studies (ACCESS). J Asthma 2010;47:257-62.
5. Peters SP, Ferguson G, Deniz Y, Reisner C. Uncontrolled asthma: a review of the prevalence, disease burden and options for treatment. Respir Med 2006;100:1139-51.
6. Szefler SJ, Wenzel S, Brown R, Erzurum SC, Fahy JV, Hamilton RG, et al. Asthma outcomes: biomarkers. J Allergy Clin Immunol 2012;129:S9-23.
7. British Thoracic Society (BTS) and the Scottish Intercollegiate Guidelines Network. British Guideline on the Management of Asthma. Thorax 2003;58:i1-94.
8. Global Initiative for Asthma. Global strategy for asthma management and prevention. Accessed July 29, 2020. https://ginasthma.org/wp-content/uploads/2016/01/GINA_Report_2015_Aug11-1.pdf
9. Davis SQ, Permutt Z, Permutt S, Naureckas ET, Bilderback AL, Rand CS, et al. Perception of airflow obstruction in patients hospitalized for acute asthma. Ann Allergy Asthma Immunol 2009;102:455-61.
10. Stahl E. Correlation between objective measures of airway calibre and clinical symptoms in asthma: a systematic review of clinical studies. Respir Med 2000;94:735-41.
11. Teeter JG, Bleecker ER. Relationship between airway obstruction and respiratory symptoms in adult asthmatics. Chest 1998;113:272-7.
12. Busse WW, Morgan WJ, Taggart V, Togias A. Asthma outcomes workshop: overview. J Allergy Clin Immunol 2012;129:S1-8.
13. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Patient-focused drug development: methods to identify what is important to patients. Accessed July 29, 2020. https://www.fda.gov/media/131230/download
14. U.S. Food and Drug Administration. Guidance for Industry: Patient-reported outcome measures: use in medical product development to support labeling claims. Accessed July 29, 2020. https://www.fda.gov/media/77832/download
15. U.S. Food and Drug Administration. Discussion document for Patient-Focused Drug Development Public Workshop: methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments Workshop date: October 15-16 2018. Accessed July 29, 2020. https://www.fda.gov/media/116277/download
16. Lasch KE, Marquis P, Vigneux M, Abetz L, Arnould B, Bayliss M, et al. PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010;19:1087-96.
17. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. Value Health 2011;14:978-88.
18. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 1—eliciting concepts for a new PRO instrument. Value Health 2011;14:967-77.
19. Krishnan JA, Lemanske RF Jr, Canino GJ, Elward KS, Kattan M, Matsui EC, et al. Asthma outcomes: symptoms. J Allergy Clin Immunol 2012;129:S124-35.
20. Nelsen LM, Gater A, Hall R, Coons SJ. Identifying and measuring the core symptoms reported by persons with asthma: a review of the existing qualitative literature and patient-reported outcome measures. C23. In: Environmental and Psychosocial Aspects of Asthma and COPD. New York, NY: American Thoracic Society; 2014. p. A4029.
21. Worth A, Hammersley V, Knibb R, Flokstra-de-Blok B, DunnGalvin A, Walker S, et al. Patient-reported outcome measures for asthma: a systematic review. NPJ Prim Care Respir Med 2014;24:1-8.
22. Gater A, Nelsen L, Fleming S, Lundy JJ, Bonner N, Hall R, et al. Assessing asthma symptoms in adolescents and adults: qualitative research supporting development of the Asthma Daily Symptom Diary. Value Health 2016;19:440-50.
23. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics

Evaluation and Research (CBER). Qualification process for drug development tools: Guidance for Industry and FDA Staff. Draft Guidance 2019. Accessed July 29, 2020. https://www.fda.gov/media/133511/download

24. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Qualification statement qualification of Asthma Daytime Symptom Diary and Asthma Nighttime Symptom Diary: patient-reported outcome instruments for measurement of symptoms of asthma. Accessed July 29, 2020. https://fda.report/media/122726/DDT+COA+006+Qualification+Statement.pdf
25. Everitt BS. Multivariate analysis: the need for data, and other problems. Br J Psychiatry 1975;126:237-40.
26. Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. Am J Respir Crit Care Med 2009;180:59-99.
27. Santanello NC, Barber BL, Reiss TF, Friedman BS, Juniper EF, Zhang J. Measurement characteristics of two asthma symptom diary scales for use in clinical trials. Eur Respir J 1997;10:646-51.
28. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol 2004;113:59-65.
29. Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika 1979;44:443-60.
30. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1-55.
31. Yu C. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Los Angeles, Calif: University of California Los Angeles; 2002.
32. Stevens JP. Applied multivariate statistics for the social sciences. Routledge; 2012.
33. Houts C, Cai L. flexMIRT user's manual version 2.0: flexible multilevel multidimensional item analysis and test scoring. Chapel Hill, NC: Vector Psychometric Group; 2013.
34. Nunnally JC. Psychometric theory 3E. New York, NY: Tata McGraw-Hill Education; 2010.
35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
36. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30-46.
37. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. Qual Life Res 2019;28:1029-33.
38. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. Health Qual Life Outcomes 2010;8:81.
39. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. Stat Med 1987;6:79-90.
40. Sullivan LM, D'Agostino RB Sr. Robustness and power of analysis of covariance applied to ordinal scaled data as arising in randomized controlled trials. Stat Med 2003;22:1317-34.
41. Mitchell J, Bradley C. Psychometric evaluation of the 12-item Well-being Questionnaire for use with people with macular disease. Qual Life Res 2001;10:465-73.
42. Cappelleri JC, Zou KH, Bushmakin AG, Alvir JM, Alemayehu D, Symonds T. Patient-reported outcomes: measurement, implementation and interpretation. Boca Raton, FL: CRC Press; 2013.
43. SAS Institute Inc. SAS Version 9. Cary, NC; 2013.
44. Muthen LK, Muthen BO. Mplus User's Guide. 7th ed. Los Angeles, CA; 2012.
45. Nelsen LM, Gater A, Bonner N, Hall R, Staunton H, Eremenco S, et al. Respondent understanding and user experience with the Asthma Daily Symptom Diary (ADSD), a novel patient-reported outcome measure. Am J Respir Crit Care Med 2017;195:A3334.