https://doi.org/10.5664/jcsm.9010

JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Psychometric evaluation of the Nocturia Sleep Quality Scale based on data from a prospective observational study

Valerie Williams, PhD[1]; Shanshan Qin, PhD[1]; Carla DeMuro Romano, MS[1]; Sandy Lewis, BSN, RN[1]; Nicole Williams, BS[1]; Stuart Yarr, MA[1]; Kristian V. Juul, PhD[2]; Fredrik L. Andersson, PhD[2]

[1]RTI Health Solutions, Research Triangle Park, North Carolina; [2]Ferring Pharmaceuticals A/S, Copenhagen, Denmark

Study Objectives: The Nocturia Sleep Quality Scale (NSQS), a novel patient-reported outcomes measure, was developed to assess the impact of sleep disturbance from nocturia. The objective of this study was to assess the psychometric properties of the NSQS, including its structure, reliability, and validity.
Methods: Data were collected in the context of a web-based, prospective, longitudinal, observational study. Participants with nocturia were randomized 1:1 to either a group that received sleep hygiene instructions, including instructions to limit liquids at nighttime and empty bladder prior to bedtime, or one that did not receive sleep instructions. All participants were asked to provide responses to the web-based questionnaires from day 1 to day 10. Psychometric analyses, aligned with current regulatory guidance, were conducted to evaluate the daily scores and 3-day average scores of NSQS items and potential composites. Item-level analyses were conducted first, followed by composite-level analyses.
Results: The NSQS items and supporting measures demonstrated very slight improvement in patient-perceived sleep disturbance from nocturia over the course of the study. NSQS test-retest reliabilities were generally satisfactory. Correlations between NSQS items and related patient-reported measures tended to support the construct validity of the NSQS, and the known-groups analyses supplied evidence of its discriminating ability. NSQS responsiveness statistics were small.
Conclusions: The NSQS is a reliable and valid measure of the impact of nocturia on patients' sleep. The present analyses lay the psychometric groundwork for the use of the NSQS in future clinical trials to support product approval and labeling claims.
Keywords: nocturia, sleep impacts, psychometric analyses, Nocturia Sleep Quality Scale (NSQS), validity, reliability
Citation: Williams V, Qin S, Romano CD, et al. Psychometric evaluation of the Nocturia Sleep Quality Scale based on data from a prospective observational study. *J Clin Sleep Med.* 2021;17(4):691–701.

BRIEF SUMMARY
Current Knowledge/Study Rationale: The Nocturia Sleep Quality Scale (NSQS) was designed to assess the impact of nocturia on sleep in a clinical trial setting. The objective of this study was to provide the first quantitative evidence in support of the psychometric properties of the NSQS, including reliability and validity, with preliminary information on responsiveness and interpretation of change.
Study Impact: The NSQS is a reliable, valid, appropriate, and useful measure of the patient-perceived impact of nocturia on sleep that was developed in accordance with the United States Food and Drug Administration's guidance for patient-focused outcome measurement. The NSQS can be used to support product approval and labeling claims in future clinical trials of nocturia treatments.

INTRODUCTION

Nocturia, by its definition, interrupts sleep through the need to void[1]; 2 or more voids per night is considered clinically significant.[2] The patient experience of nocturia leads to bothersome sleep disturbance and quantifiable effects on a person's physical, emotional, and social well-being, including excessive daytime sleepiness and reductions in cognitive performance.[3–6] Furthermore, getting up at night to void also creates a substantial impact on work engagement, work productivity, and employment.[7] The quality-of-life burden on patients from nocturia has been associated with most health-related quality-of-life dimensions,[2] and these burdens extend across cultures and sexes.[8] Moreover, the impact of nocturia on the sleep quality of elderly patients may be much greater than previously understood and could contribute to an increased fall risk.[9–11]

A variety of patient-reported outcome (PRO) instruments are available to assess the impact of sleep disturbance in general or in patients with lower urinary tract symptoms. The Pittsburgh Sleep Quality Index evaluates sleep quality and disturbances over a 1-month time period.[12] The American Urological Association Symptom Index/International Prostate Symptom Score (AUA-SI/IPSS) and the Nocturia, Nocturnal Enuresis, and Sleep-interruption Questionnaire (NNES-Q) address sleep disturbance in a more restricted lower urinary tract symptoms context and with a focus on nocturnal voiding, respectively.[13,14] However, these measures are not specific to nocturia or were not specifically developed in accordance with the US Food and Drug Administration's guidance for patient-focused outcome measurement.[15–18] These guidance documents stipulate that PRO measures referenced in product labeling must accurately reflect the patient experience and be developed using extensive input from patients in the population of interest and based on appropriate methodology

Table 1—Inclusion and exclusion criteria.

Inclusion Criteria
Aged ≥ 18 years
Experienced self-reported nocturia (≥ 2 nocturnal voids per night) for ≥ 6 months
Experienced sleep disturbance due to nocturia, as defined by response ≥ 2 to the following question:
How much has your sleep been disturbed by having to get up to urinate during the night during the past 2 weeks?
0 = Not at all
1 = Slightly
2 = Moderately
3 = Quite a bit
4 = Extremely
Willing and able to provide informed consent
Able to read and speak English
Exclusion Criteria
Self-report of any of the following conditions:
Sleep apnea
Severe urinary incontinence (eg, requires nighttime protective garments)
Poorly controlled type 1 or type 2 diabetes mellitus (eg, condition is untreated or has required a change or adjustment in medications in the past 30 days)
Poorly controlled hypertension (eg, condition is untreated or has required a change or adjustment in medications in the past 30 days)
Heart failure
Current pregnancy
Edema in the past 2 weeks
Self-reported polydipsia (excessive fluid intake, exceeding 12 cups [3 L]/24 hours), or feeling extremely thirsty and drinking a lot of fluids unrelated to exercise, in the past 2 weeks
Urinary tract infection in the past 2 weeks
Employed as a shift worker
Presence of cognitive impairment, clinical dementia, or psychiatric illness or other neurodegenerative disease (eg, Parkinson disease, Alzheimer disease) significant in the opinion of the investigator and would preclude participation in the study

that has been thoroughly validated in the population expected to participate in the clinical trials. To fill the gap and enrich the literature, the Nocturia Sleep Quality Scale (NSQS) was developed and refined on the basis of literature review, concept elicitation, and cognitive debriefing interviews with patients and consultation with clinical experts.[8] The NSQS was designed to assess the impact of nocturia on sleep in a clinical trial setting. Previous cognitive debriefing interviews conducted with patients who had nocturia have confirmed the comprehensiveness of the NSQS and its relevance, providing support for the content validity and ability of items to reflect patient perception of important, meaningful, and relevant nocturia-related sleep impact.[8]

The objective of this study was to provide the first quantitative evidence in support of the psychometric properties of the NSQS, including reliability and validity, with preliminary information on responsiveness and interpretation of change.

METHODS

This study used a prospective, observational, longitudinal design that was reviewed on ethical grounds by the RTI International institutional review board and granted an exemption from full review. Individuals with nocturia were recruited nationally through two qualitative research facilities (located near Philadelphia, Pennsylvania, and Raleigh, North Carolina). The qualitative research staff notified potential participants about the study by posting a recruitment advertisement on a website or by e-mails sent to members of their participant database. After being screened according to inclusion and exclusion criteria (**Table 1**), eligible individuals were contacted by trained staff at the qualitative research facility to review the informed consent form and obtain each individual's agreement to participate in the study. Participants were randomized 1:1 to a group that received sleep hygiene instructions or to a group that did not receive sleep hygiene instructions. Participants in the group receiving sleep hygiene instructions were provided with lifestyle/behavioral change guidance (in the supplemental material). Specifically, they were directed to not drink water or liquids (in particular, caffeinated and alcoholic drinks) within 1.5 hours of bedtime, to elevate their legs before going to bed to help redistribute fluids back into the bloodstream and reduce the need to urinate, and to empty their bladders within 15 minutes of going to bed with the intention to sleep. Participants were expected to use their own best judgment about

Table 2—Schedule of key events for NSQS validation analyses.

Procedure	Screening (Day 0)	Online Survey (Day 1)	Online Survey (Days 2–10)
Demographics	✓		
Medical history	✓		
NSQS		✓	✓
Number of nighttime voids	✓	✓	✓
PGI-S		✓	✓ (days 7 and 10)
PGI-I			✓ (days 7 and 10)
NID		✓	✓ (days 7 and 10)

NID = Nocturia Impact Diary, NSQS = Nocturia Sleep Quality Scale, PGI-I = Patient Global Impression of Improvement, PGI-S = Patient Global Impression of Severity.

duration for leg elevation. The advice to elevate legs before going to bed[19,20] is supported by evidence that leg edema is a cause of nocturnal polyuria and nocturia through resorption of fluid when supine.[21,22] While there are other more specific means to reduce leg edema (eg, compression stockings), participants were restricted to benign sleep hygiene behaviors recommended by the Urology Care Foundation[23] as this was not an interventional study. All participants in both groups were asked to provide responses to the web-based clinical outcomes assessment (COA) measures (**Table 2**) at home each morning upon awakening from day 1 to day 10. Participants received a nominal amount of compensation for their time to complete the daily diary for 10 days.

Measures

The NSQS is a six-item, patient-reported measure of the impact of nocturia on nighttime sleep quality and is intended to assess change in the impact of nocturia after treatment in a standardized manner.[8] The COA measures used to evaluate the psychometric performance of the NSQS included the number of nighttime voids, the Patient Global Impression of Severity (PGI-S), the Patient Global Impression of Improvement (PGI-I), and the Nocturnal Impact Diary (NID). For the number of nighttime voids, patients were asked to recall each morning the number of times they urinated during the night using a value ranging from 0 to 10 and to not include their first void of the morning upon awakening for the day. Participants provided an overall assessment of the severity of their nocturia symptoms using the PGI-S on days 1, 7, and 10. The second global item, the PGI-I, asked participants on days 7 and 10 to provide an overall retrospective assessment of the change in their nocturia symptoms since the start of the study. The NID[24] is a 12-item PRO instrument designed to assess the daily impact of nocturia on the everyday life of patients and was completed by all participants on days 1, 7, and 10 in the afternoon or evening. The first 11 NID items are summed for a total score that is transformed to a 0–100 scale, whereas question 12 separately refers to the overall impact of nocturia.

Analytic methods

Psychometric analyses, aligned with current regulatory guidance,[15,17] were used to evaluate the daily scores and 3-day average scores (in alignment with clinical trial practices for nocturia diaries)[25–28] of NSQS items and potential composite scores. Item-level analyses were conducted first to evaluate the behavior of the individual NSQS items, followed by parallel composite-level analyses to describe and assess the properties of potential scores based on groups of NSQS items. Most analyses were performed using SAS version 9.4 or higher (SAS Institute Inc., Cary, NC, 2012). Mplus version 7.4 (Muthén and Muthén, Los Angeles, CA, 1998–2015) was used to conduct the factor analyses. All statistical tests were 2-tailed, and, unless otherwise stated, a type I error rate of 1% was applied to each individual hypothesis test. Rather than relying on significance tests alone, we emphasized interpretation of patterns of results and effect-size estimates in the psychometric analyses. PRO measures used in the validation of the NSQS were scored according to the instrument developers' guidelines (including missing data rules). Because of the design of the web-based survey, there were no missing item-level responses to the questionnaires. No imputation was used when computing NSQS composite scores or the 3-day averages, which required complete item and daily scores. Demographic and clinical characteristics were tabulated at baseline, as well as descriptive statistics of all supporting measures.

Descriptive statistics for each NSQS item were computed to characterize the sleep quality of the study sample. Response frequency distributions for each NSQS item were tabulated to characterize the behavior of each item and assess floor and ceiling effects and other potential response biases. A floor or ceiling effect would require that at least 40% (for 5-point items) or 33.34% (for 6-point items) of the participants select the worst or best response category, respectively. For each item, weighted kappa coefficients[29] of daily item scores and intraclass correlation coefficients (ICCs) of 3-day item averages were computed to assess item-level test-retest reliability in patients who rated themselves as unchanged over the time interval of interest. The Fleiss–Cohen quadratically weighted kappa coefficient evaluates the strength of agreement between 2 measurements while accounting for the levels of differences in ranked categories; it is equivalent to the ICC used for continuous outcomes.[30] Kappa statistics can range from −1 to 1 and can be interpreted such that ≤ 0 is poor, 0–0.2 indicates slight agreement, 0.21–0.4 indicates fair agreement, 0.41–0.6

indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–1.00 indicates almost perfect agreement.[31] A 2-way mixed effects (random subjects × fixed time) analysis of variance (ANOVA) for absolute agreement between single measures was used to compute ICC estimates of test-retest reliability.[32] It is generally recommended that ICCs be at least 0.70 for multi-item scales.[33] In the first test-retest analysis, day 9 NSQS scores were the "test" data and day 10 NSQS scores were the "retest" data. This analysis included only those participants whose PGI-I ratings at day 10 equaled no change. A second test-retest analysis used the day 1–2–3 average as "test" and the day 8–9–10 average as "retest," and it included only those participants whose PGI-S ratings at day 1 were equal to their PGI-S ratings at day 10.

Correlational analyses between NSQS items and composite scores and supporting COA measures were conducted to examine the cross-sectional and longitudinal construct validity of the NSQS. In all construct validity applications, hypotheses are proposed regarding the relationships between the constructs of the COA of interest and available external measures. These hypotheses should indicate the direction (positive or negative) of the relationships and the magnitude (either in strength or relative to other planned correlations).[34] The goal was to demonstrate stronger association among measures addressing similar constructs (convergent validity) in comparison to measures addressing more disparate constructs. Cohen's[35] guidelines for the interpretation of correlation coefficients describe correlations ≥ 0.50 as large or strong, correlations of 0.30 to 0.49 as medium or moderate, correlations of 0.10 to 0.29 as small, and correlations < 0.10 as trivial.

The ability of the NSQS to discriminate between subgroups based on empirical knowledge of differences was assessed by using ANOVAs based on a priori hypotheses to examine mean differences in NSQS scores between participants classified into known groups; for example, it was hypothesized that participants who were given sleep hygiene instructions or who reported fewer than 2 voids per night would have better (lower) NSQS scores compared with participants who did not receive sleep instructions or who reported 2 or more voids per night.

Responsiveness was evaluated by computing paired *t* tests and effect-size estimates of change (ie, the mean change from baseline to the end of study divided by the standard deviation [SD] at baseline)[36,37] for each NSQS score. Cohen[35] provides a general rule of thumb for the interpretation of effect size estimates in SD units: values of approximately 0.20 represent small effects, those of approximately 0.50 represent moderate effects, and those greater than approximately 0.80 represent large effects.

To evaluate the NSQS structure, interitem correlations were computed, with special attention given to items with very large correlation coefficients (> |0.80|), which were flagged for possible redundancy. Exploratory factor analyses were performed on day 1–2–3 average item scores by using maximum likelihood estimation with robust standard errors; varying numbers of factors were extracted with oblique quartimin rotation. Based on the exploratory factor analyses results, confirmatory factor analyses (CFAs) were conducted on day 8–9–10 item scores, also with the use of maximum likelihood estimation with robust standard errors . The 2-factor CFA model allowed the factors to be correlated, by which each factor mean was fixed to 0 and each factor variance was fixed to 1 for model identification. Cronbach's coefficient alpha[38] was computed to describe the internal consistency of NSQS item subsets. The approximate range of optimal Cronbach's alphas is between 0.70 and 0.90.[29] In addition to the qualitative data from the NSQS development,[8] the results of the interitem correlations, factor analyses, and internal consistency reliabilities provided guidance regarding the optimal scoring of the NSQS. Candidate scoring algorithms were evaluated with respect to reliability, validity, and responsiveness, as described above.

To facilitate interpretation of change, preliminary responder thresholds were estimated using anchor-based methods, with the PGI-S as the primary anchor variable; a participant was classified as a responder if he or she reported a 1-unit or more improvement from day 1 to day 10. The PGI-I was used as a secondary anchor variable. Distribution-based methods (half-SD and the standard error of measurement) were also applied. Empirical cumulative distribution function and probability density function plots were developed to aid in the understanding and interpretation of estimated responder thresholds.

RESULTS

Participant characteristics

Table 3 presents the participant characteristics at baseline. Of the 100 participants, 50 (50%) were male and 50 (50%) were female. The mean age of the sample was 50.39 years (SD = 14.1 years), with a median age of 49.0 (range, 22–77 years). The mean age of female participants was 48.3 (SD = 14.4), and the mean age of males was 52.5 (SD = 13.5); this difference was not statistically significant ($F = 2.28$, $P = .1339$). Six participants (6%) reported that they had an enlarged prostate (or benign prostatic hyperplasia) and 6 participants (6%) reported that they had overactive bladder, in addition to nocturia. There were no statistical differences ($\alpha = 0.01$) at baseline between the sleep hygiene instructions group and the no sleep instructions group in terms of any of the characteristics displayed in **Table 3** or with respect to demographics or medical history data collected (eg, highest education level, history of enlarged prostate, or overactive bladder).

NSQS item-level results

There was a slight improvement from day 1 to day 10 and from day 1–2–3 to day 8–9–10 in the NSQS item scores (**Table 4**). Very slight improvements were also observed in the supporting measures over the course of the 10-day study (**Table 4**). Item-level response frequency distributions showed no indication of floor or ceiling effects and supported the appropriateness of the NSQS item response categories. Missing data were not problematic, and there was no pattern of attrition or fatigue. The percentage of missing data ranged from to 5% on day 10 to 13% on day 7, and the percentage of missing data was the same across items because of the design of the web-based survey.

Using data from the subgroup of participants with the same PGI-S ratings at day 1 and day 10 (n = 41), the test-retest

Table 3—Patient characteristics at screening.

Characteristic	Overall (n = 100)	
Age, mean (SD), median, minimum–maximum, years	50.39 (14.1), 49.0, 22–77	
Sex, n (%)		
Male	50 (50.0)	
Female	50 (50.0)	
Race/ethnicity, n (%)		
White	64 (64.0)	
Black or African American	19 (19.0)	
American Indian or Alaska native	1 (1.0)	
Asian	3 (3.0)	
Native Hawaiian or other Pacific Islander	0 (0.0)	
Other	13 (13.0)	
Hispanic or Latino	14 (14.0)	
Usual number of nighttime awakenings because of the need to urinate—over the past 6 months [a]		
2 times per night	53 (53.0)	
3 times per night	29 (29.0)	
4 or more times per night	18 (18.0)	
Number of nighttime voids—over the past 2 weeks[a]		
2 times per night	50 (50.0)	
3 times per night	32 (32.0)	
4 or more times per night	18 (18.0)	
How much has your sleep been disturbed by having to get up to urinate during the night in the past 2 weeks?		
Moderately	31 (31.0)	
Quite a bit	44 (44.0)	
Extremely	25 (25.0)	
PGI-S, n (%)	**Day 1**	**Day 10**
1 = None	0 (0.0)	0 (0.0)
2 = Mild	23 (23.0)	43 (45.3)
3 = Moderate	61 (61.0)	41 (43.2)
4 = Severe	16 (16.0)	11 (11.6)

PGI-S = Patient Global Impression of Severity, "Check the one number that best describes how your nighttime urination is now (based on the past 24 hours)", SD = standard deviation. [a]As recalled by the participant at screening.

ICCs for the day 1–2–3 to day 8–9–10 averages ranged from 0.58 for Item 4 (How restful?) to 0.84 for Item 1 (How many times awakened?). The weighted kappa test-retest reliabilities based on day 9 to day 10 ranged from 0.38 for item 4, indicating fair agreement,[31] to 0.75 for item 1, indicating substantial agreement.

The hypothesized correlations between NSQS item scores and supporting measures followed the expected patterns (**Table S1**), eg, strong correlations were observed between NSQS item 2 (Total time awake?) and NID item 7 (Lay awake) (r = .56 for the day 1–2–3 average to r = .80 for day 10). As expected, the strongest correlations were between NSQS item 1 (How many times awakened?) and the closely related number of nighttime voids (r = .84 for the day 1–2–3 average and r = .87 for the day 8–9–10 average). The correlations between change in NSQS item 1 and change in number of nighttime voids (r = .76 for day 1 to day 10 and r = .63 for day 1–2–3 to day 8–9–10 average) were also very strong (**Table S2**). Effect-size estimates of change ranged from −0.09 (change from day 1–2–3 to day 8–9–10 for item 4) to –0.38 (change from day 1–2–3 to day 8–9–10 for item 5), generally small in magnitude. The test statistic for item 5 (How tired?) change from day 1–2–3 to day 8–9–10 was statistically significant, as was that for item 1 (How many times awakened?) change from day 1 to day 10.

NSQS structure

All interitem correlations were positive in sign, and the majority were statistically different from 0 (P = .01) and at least moderate ($|r| \geq .30$) in size (**Table S3**). The 2-factor exploratory factor

Table 4—Descriptive statistics for NSQS items and supporting COA measures.

COA, Mean (SD), Median	Day 1 (n = 100)	Day 10 (n = 95)	Day 1–2–3 Average (n = 89)	Day 8–9–10 Average (n = 82)	Change from Day 1 to Day 10 (n = 95)	Change from Day 1–2–3 to Day 8–9–10 (n = 76)
NSQS						
1. How many times awakened?	3.20 (1.10), 2.5	2.79 (1.26), 2.5	3.00 (1.05), 2.92	2.84 (1.20), 2.5	−0.39 (1.08), 0.0	−0.18 (0.77), 0.0
2. Total time awake?	2.09 (1.04), 2.0	1.96 (0.99), 2.0	2.05 (0.92), 2.0	1.91 (0.90), 1.67	−0.16 (0.91), 0.0	−0.09 (0.62), 0.0
3. Get up earlier?	2.10 (1.45), 2.0	1.81 (1.32), 2.0	1.90 (1.11), 2.0	1.71 (1.05), 1.67	−0.33 (1.46), 0.0	−0.17 (0.91), 0.0
4. How restful?	3.16 (1.06), 2.5	3.12 (1.21), 3.75	3.10 (0.86), 2.92	3.03 (0.98), 2.92	−0.11 (1.38), 0.0	−0.08 (0.94), 0.0
5. How tired?	2.59 (1.21), 2.5	2.22 (1.32), 2.5	2.56 (1.12), 2.5	2.18 (1.09), 2.08	−0.33 (1.57), 0.0	−0.42 (1.01), −0.42
6. Overall sleep quality?	2.73 (0.98), 2.5	2.63 (1.12), 2.5	2.73 (0.92), 2.5	2.57 (0.92), 2.5	−0.12 (1.14), 0.0	−0.16 (0.72), 0.0
Number of Nighttime voids	2.86 (1.26), 3.0	2.56 (1.51), 2.0	2.75 (1.25), 2.7	2.61 (1.48), 2.0	−0.34 (0.99), 0.0	−0.20 (0.79), 0.0
PGI-S	2.93 (0.62), 3.0	2.66 (0.68), 3.0	—	—	−0.25 (0.67), 0.0	—
PGI-I	—	3.75 (0.79), 4.0	—	—	—	—
NID total score	39.44 (17.81), 34.1 (n = 88)	36.95 (21.10), 35.2 (n = 82)	—	—	−1.36 (15.73), −2.3 (n = 75)	—

NID = Nocturia Impact Diary, NSQS = Nocturia Sleep Quality Scale, PGI-I = Patient Global Impression of Improvement, PGI-S = Patient Global Impression of Severity, SD = standard deviation. Note: The 5-point response scales for items 1, 4, 5, and 6 were rescaled to (0, 1.25, 2.5, 3.75, 5). NSQS Item 4 (How restful?) is reverse-scored such that 0 = "Extremely restful" and 5 = "Not at all restful.

Table 5—NSQS factor analysis factor loadings (standard errors) and model fit.

NSQS Item	1-Factor Solution	2-Factor Solution Factor 1	2-Factor Solution Factor 2
EFA: Day 1–2–3 average (n = 89)			
1. How many times awakened?	0.57* (0.08)	0.25 (0.26)	0.39 (0.25)
2. Total time awake?	0.65* (0.07)	0.98* (0.28)	−0.02 (0.05)
3. Get up earlier?	0.59* (0.08)	0.47 (0.40)	0.26 (0.40)
4. How restful?	0.81* (0.05)	−0.07 (0.09)	0.86* (0.08)
5. How tired?	0.97* (0.02)	0.03 (0.07)	0.85* (0.06)
6. Overall sleep quality?	0.96* (0.02)	0.03 (0.12)	0.95* (0.10)
RMSEA, SRMR	0.133, 0.055	0.000, 0.011	
CFI, TLI	0.952, 0.920	1.000, 1.020	
CFA: Day 8–9–10 average (n = 82)			
1. How many times awakened?	0.62* (0.07)	0.67* (0.06)	—
2. Total time awake?	0.72* (0.07)	0.95* (0.03)	—
3. Get up earlier?	0.73* (0.07)	0.91* (0.04)	—
4. How restful?	0.88* (0.03)	—	0.87* (0.03)
5. How tired?	0.90* (0.03)	—	0.90* (0.03)
6. Overall sleep quality?	0.95* (0.02)	—	0.98* (0.02)
RMSEA, SRMR	0.322, 0.082	0.000, 0.029	
CFI, TLI	0.821, 0.702	1.000, 1.004	

CFA = confirmatory factor analysis, CFI = comparative fit index, EFA = exploratory factor analysis, NSQS = Nocturia Sleep Quality Scale, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, TLI = Tucker-Lewis index. *$P < 0.05$ for H_0: loading = 0.

analyses solution showed excellent fit (**Table 5**), with items 4, 5, and 6 loading strongly (≥ .85) on the second factor, and items 2 and 3 loading moderately to strongly (.98 and .47, respectively) on the first factor. Item 1 obtained a stronger loading (.39) on the second factor compared with the loading on the first factor (.25); however, the standard errors were relatively large, indicating uncertainty. The interfactor correlation for the 2-factor model was $r = .64$. In the best-fitting 2-factor CFA model, items 1, 2, and 3 loaded exclusively on the first factor and items 4, 5, and 6 loaded on the second factor, with a negative residual covariance between item 2 and item 6. The interfactor correlation for the 2-factor CFA model was $r = .75$. It was tentatively decided that

Table 6—Psychometric properties of NSQS composite scores.

Time Point	Total Average Score	Lost Sleep Time Average Score	Impacts Average Score
Mean (SD), median			
Day 1 (n = 100)	2.64 (0.83), 2.54	2.46 (0.90), 2.50	2.83 (0.91), 2.71
Day 1–2–3 Average (n = 89)	2.56 (0.79), 2.39	2.32 (0.83), 2.22	2.80 (0.89), 2.78
Day 8–9–10 Average (n = 82)	2.37 (0.87), 2.28	2.15 (0.93), 2.00	2.59 (0.94), 2.57
Change from day 1–2–3 to day 8–9–10 (n = 76)	−0.18 (0.61), −0.14	−0.15 (0.60), −0.14	−0.22 (0.77), −0.14
Cronbach's Alpha			
Day 1–2–3 Average (n = 89)	0.88	0.73	0.91
Day 8–9–10 Average (n = 82)	0.92	0.86	0.93
Test-retest ICC (95% CI)			
Day 9 to Day 10 (n = 56)	0.58 (0.37−0.73)	0.68 (0.51−0.80)	0.45 (0.21−0.64)
Day 1–2–3 Average to Day 8–9–10 Average (n = 41)	0.80 (0.60−0.90)	0.84 (0.70−0.92)	0.73 (0.50−0.85)
Construct validity correlations — Day 1–2–3 average, Day 8–9–10 average			
Number of Nighttime Voids	0.65*, 0.72*	0.68*, 0.78*	0.50*, 0.56*
PGI-S	0.66*, 0.65*	0.63*, 0.63*	0.57*, 0.57*
NID Total	0.74*, 0.77*	0.67*, 0.73*	0.68*, 0.68*
Construct Validity Correlations of Change — Day 1 to Day 10 (n = 75 to 95), Day 1–2–3 average to Day 8–9–10 average (n = 64 to 76)			
Number of Nighttime Voids	0.56*, 0.46*	0.53*, 0.48*	0.48*, 0.35*
PGI-S	0.37*, 0.14	0.38*, 0.16	0.30*, 0.11
PGI-I	0.35*, 0.26	0.38*, 0.27	0.27, 0.20
NID Total	0.48*, 0.32	0.37*, 0.28	0.48*, 0.27
Responsiveness — Effect-size estimate, paired *t* (*P* value)			
Day 1 to day 10 (n = 95)	−0.29, 2.61 (0.0105)	−0.32, 3.41 (0.0010)	−0.20, 1.56 (0.1222)
Day 1–2–3 to day 8–9–10 (n = 76)	−0.23, 2.64 (0.0100)	−0.18, 2.17 (0.0335)	−0.25, 2.49 (0.0150)

NID = Nocturia Impact Diary, NSQS = Nocturia Sleep Quality Scale, PGI-I = Patient Global Impression of Improvement, PGI-S = Patient Global Impression of Severity, SD = standard deviation. * $P < 0.01$ for H_0: $\rho = 0$.

the first factor described lost sleep time and the second factor described impacts.

NSQS composite-level results

In addition to a total NSQS score, both lost sleep time (items 1, 2, 3) and impacts (items 4, 5, 6) were further evaluated (**Table 6**). Similar to the observed average changes in NSQS items and the other COAs included in the study, there were very slight improvements in all composite scores across time. Cronbach's alphas for all NSQS composites exceeded 0.70 at all time points. The test-retest reliabilities for NSQS composites using the day 1–2–3 to day 8–9–10 data were all greater than 0.70; the test-retest reliabilities using the day 9 to day 10 data were somewhat lower.

All correlations between the 3 NSQS composite scores and the other measures in the study were at least moderate in size. As expected, the correlations were relatively strong between changes in NSQS composite scores and changes in number of nighttime voids (**Table 6**); the correlation between change in the NSQS total score and change in the NID total score was moderate, $r = .32$. The correlations between NSQS change scores and PGI-S change and PGI-I ratings were relatively small when day 1–2–3 to day 8–9–10 data were used; however, the correlations were moderate, using change from day 1 to day 10. In the first set of known-groups analyses of variance (**Table S4**), participants in the sleep hygiene instructions group achieved better NSQS composite scores compared with participants in the no sleep instructions group; however, none of the differences were statistically significant. In the second known-groups analysis (**Table S5**), participants who reported fewer than 2 nighttime voids achieved better NSQS composite scores compared with participants who reported 2 or more nighttime voids; these differences were all statistically significant ($P < .05$). With respect to the responsiveness of the NSQS, effect-size estimates of change were generally small (**Table 6**).

Preliminary responder thresholds were estimated using the PGI-S. Although it was expected that the correlations between changes in anchor variables and changes in NSQS scores would be greater than 0.30,[39] the correlations based on change from day 1–2–3 to day 8–9–10 were lower: $r = .14$ between PGI-S change and NSQS total change, $r = .16$ between PGI-S change and lost sleep time change, and $r = .11$ between PGI-S change and impacts change (**Table 6**). The correlations based on the PGI-I were somewhat stronger: $r = .26$ between the PGI-I and NSQS total change, $r = .27$ between the PGI-I and lost sleep time change, and $r = .20$ between the PGI-I and impacts change.

Table 7—Interpretation of change, day 1–2–3 average to day 8–9–10 average.

Responder Definition Estimation Method	Total Average Score	Lost Sleep Time Average Score	Impacts Average Score
PGI-S Anchor Change — mean, median, mode			
2-point improvement (n = 2)	–0.38, –0.38, no mode	–0.07, –0.07, no mode	–0.69, –0.69, no mode
1-point improvement (n = 25)	–0.18, –0.25, mm	–0.20, –0.17, mm	–0.16, –0.14, –0.83
No difference (n = 41)	–0.27, –0.21, 0.47	–0.22, –0.22, 0.25	–0.32, –0.42, –0.56
1-point worsening (n = 8)	0.29, 0.37, no mode	0.34, 0.37, 0.39	0.24, 0.42, mm
PGI-I Anchor — mean, median, mode			
1 = very much better (n = 1)	−1.15, −1.15, −1.15	−0.36, −0.36, −0.36	−1.94, −1.94, −1.94
2 = much better (n = 1)	−0.36, −0.36, −0.36	−0.72, −0.72, −0.72	−0, −0, 0
3 = a little better (n = 21)	−0.25, −0.25, mm	−0.31, −0.36, −0.22	−0.19, 0, 0
4 = no change (n = 47)	−0.19, −0.14, 0.47	−0.10, 0.11, 0.25	−0.27, −0.42, −0.56
5 = a little worse (n = 4)	0.30, 0.26, no mode	0.24, 0.31, no mode	0.35, 0.21, no mode
6 = much worse (n = 1)	0.33, 0.33, 0.33	0.39, 0.39, 0.39	0.28, 0.28, 0.28
7 = very much worse (n = 1)	0.11, 0.11, 0.11	−0.33, −0.33, −0.33	0.56, 0.56, 0.56
Half-SD	0.39	0.42	0.45
SEM	0.35	0.33	0.47

mm = multiple modes, NSQS = Nocturia Sleep Quality Scale, PGI-I = Patient Global Impression of Improvement, PGI-S = Patient Global Impression of Severity, SD = standard deviation, SEM = standard error of measurement. Note: No patient improved by more than 2 points on the PGI-S.

For completeness, responder definitions were estimated with both anchor variables (**Table 7**). Using the primary anchor of a 1-point improvement on the PGI-S item, the responder definition for day 1–2–3 to day 8–9–10 change in NSQS total scores is –0.18 point (n = 25)—that is, in the present context, a decrease of 0.18 points on the NSQS total score indicates improvement. Using the secondary anchor of "much better" on the PGI-I, the responder definition for the day 1–2–3 to day 8–9–10 changes in NSQS total scores is –0.36 of a point (n = 1). Using "a little better" on the PGI-I, the responder definition for day 1–2–3 to day 8–9–10 changes in NSQS total scores is –0.25 (n = 21). The half-SDs and standard errors of the measurement resulted in reasonably consistent estimates. **Figure 1** depicts the cumulative distribution functions and probability density functions for change in NSQS total scores from day 1–2–3 to day 8–9–10. Although the 4 functions overlap and cross repeatedly, the 1-point worsening function is mostly to the right of the other functions (top panel); the probability density functions are similarly overlapping (bottom panel).

DISCUSSION

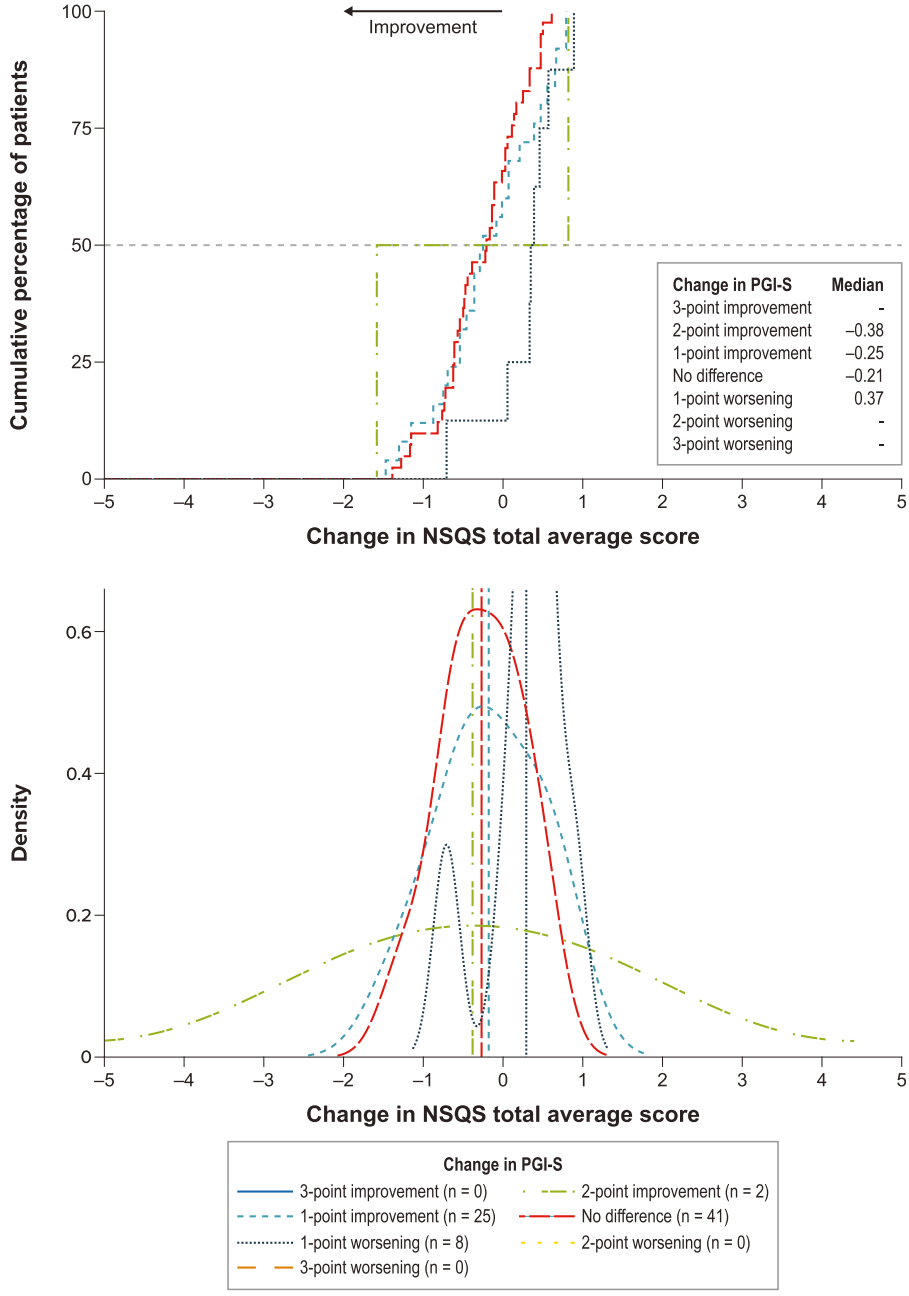
This study provides important information regarding the structure, behavior, properties, and interpretation of the NSQS. The NSQS items and composite scores, as well as the other available PRO measures, indicated small improvements in the patient-perceived impacts of nocturia over the course of the 10-day observational study. Item-level descriptive statistics and frequency distributions supported the appropriateness of the NSQS response categories with no distributional anomalies such as ceiling or floor effects that would impair the NSQS's responsiveness to change in a clinical trial. Item-level test-retest reliabilities were satisfactory, although somewhat low for item 4 (How restful?). Correlations between the NSQS items and other related patient-reported measures tended to follow the expected patterns, providing support for construct validity, and the known-groups analyses supplied evidence of the discriminating ability of the NSQS items. Effect-size estimates were generally small, as would be expected based on the very small improvements in the patient-perceived impacts of nocturia over the course of the study.

Taking into account the item-level results and the factor analyses, as well as the qualitative research conducted during the development of the NSQS, the 6-item NSQS total score and 2 additional composite scores, lost sleep time (items 1, 2, and 3) and impacts (items 4, 5, and 6), emerged as measures potentially capable of supporting study endpoints. The internal consistency reliabilities of the NSQS composite scores were highly satisfactory, as were the composite-level test-retest reliabilities, construct validity correlations, and known-groups analyses of variance. NSQS composite-level responsiveness statistics were small, again reflecting the only slight improvements in the patient-reported impact of nocturia during the 10-day observational study.

Although the establishment of a responder definition occurs as a process over multiple assessments and across a wide range of studies, various methods were used to explore potential responder definitions for the NSQS composite scores. A preliminary working value for responder thresholds defining meaningful change on the NSQS composite scores in a population of adults with nocturia is approximately 0.5 points on the 0 to 5 NSQS score scale.

One important limitation of the present research is that only very small changes in the NSQS were observed during the course of the 10-day observational study. In the context of a

Figure 1—Cumulative distribution function of change in NSQS total scores and probability density function of change.



Cumulative distribution function of change in NSQS total scores from day 1–2–3 to day 8–9–10 (top panel), and probability density function of change (bottom panel), by PGI-S change. NSQS = Nocturia Sleep Quality Scale; PGI-S = Patient Global Impression of Severity.

chronic disease, a 10-day time period is very short and limits the assessment of key longitudinal psychometric properties, in particular, responsiveness to change and the definition of responder thresholds. However, the limited follow-up period was necessary to obtain data for the initial psychometric evaluation of the NSQS. Future research with a larger sample, a longer time period, and greater improvements in nocturia is necessary to confirm the NSQS's responsiveness to change and define responder thresholds. Another limitation of this study was the use of United States-based participants only—despite the intended goal of creating items to be culturally generic, the NSQS will require cross-cultural validation in languages other than US English.

CONCLUSIONS

The NSQS is a reliable, valid, appropriate, and useful measure of the patient-perceived impact of nocturia on sleep. Although the nighttime sleep impacts are considered the most

bothersome impact of polyuria,[40] until now there has not been a measure specific to nocturia and also developed in accordance with the Food and Drug Administration PRO guidance. The NSQS fills this gap as an essential tool to support product approval and labeling claims in future clinical trials of nocturia treatments.

ABBREVIATIONS

ANOVA, analysis of variance
CDF, cumulative distribution function
CFA, confirmatory factor analyses
ICC, intraclass correlation coefficients
NID, Nocturnal Impact Diary
NSQS, Nocturia Sleep Quality Scare
PGI-I, Patient Global Impression of Improvement
PGI-S, Patient Global Impression of Severity
PRO, patient-reported outcome
SD, standard deviation

REFERENCES

1. van Kerrebroeck P, Abrams P, Chaikin D, et al. Standardisation Sub-committee of the International Continence Society. The standardisation of terminology in nocturia: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn*. 2002;21(2):179–183.
2. Tikkinen KA, Johnson TM 2nd, Tammela TL, et al. Nocturia frequency, bother, and quality of life: how often is too often? A population-based study in Finland. *Eur Urol.* 2010;57(3):488–496.
3. Ancoli-Israel S, Bliwise DL, Nørgaard JP. The effect of nocturia on sleep. *Sleep Med Rev.* 2011;15(2):91–97.
4. Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res.* 2003;12(1):1–12.
5. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*. 2003;26(2):117–126.
6. Banks S, Dinges DF. Behavioral and physiological consequences of sleep restriction. *J Clin Sleep Med.* 2007;3(5):519–528.
7. Zeng VY, Milligan G, Piercy J, Anderson P, Andersson FL. Impact of nocturia on patients' health-related quality of life and healthcare resource utilisation compared with OAB and BPH: Results from an observational survey in European and American patients. *Int J Clin Pract.* 2019; Aug 27:e13408.
8. Romano CD, Lewis S, Barrett A, et al. Development of the Nocturia Sleep Quality Scale: a patient-reported outcome measure of sleep impact related to nocturia. *Sleep Med.* 2019;59:101–106.
9. Bliwise DL, Foley DJ, Vitiello MV, Ansari FP, Ancoli-Israel S, Walsh JK. Nocturia and disturbed sleep in the elderly. *Sleep Med.* 2009;10(5):540–548.
10. Hafner M, Pollard J, Troxel W, et al. *How Frequent Night-Time Bathroom Visits Can Negatively Impact Sleep, Well-Being and Productivity: Examining The Associations Between Nocturia, Well-Being and Economic Outcomes in a Working-Age Population*. Santa Monica, CA: RAND Corporation; 2019, https://www.rand.org/pubs/research_reports/RR3043.html
11. Weidlich D, Andersson FL, Oelke M, Drake MJ, Jonasson AF, Guest JF. Annual direct and indirect costs attributable to nocturia in Germany, Sweden, and the UK. *Eur J Health Econ.* 2017;18(6):761–771.
12. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193–213.
13. Barry MJ, Fowler FJ Jr, O'leary MP, et al; Measurement Committee of the American Urological Association. The American Urological Association Symptom Index for benign prostatic hyperplasia. *J Urol.* 2017197, 2S, 2s:S189–S197.
14. Bing MH, Moller LA, Jennum P, Mortensen S, Lose G. Validity and reliability of a questionnaire for evaluating nocturia, nocturnal enuresis and sleep-interruptions in an elderly population. *Eur Urol.* 2006;49(4):710–719.
15. Food and Drug Administration. 2009. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. https://www.fda.gov/media/77832/download. Accessed May 16, 2019.
16. Food and Drug Administration. 2013. Roadmap to patient-focused outcome measurement in clinical trials. https://www.fda.gov/media/87004/download. Accessed May 15, 2019.
17. Food and Drug Administration. 2018. Patient-focused drug development (PFDD) guidance: methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. Attachment to guidance 3 discussion document – appendices. https://www.fda.gov/media/116281/download. Accessed June 26, 2019.
18. Food and Drug Administration. 2019. Patient-focused drug development: methods to identify what is important to patients. Guidance for industry, Food and Drug Administration staff, and other stakeholders. Draft Guidance. https://www.fda.gov/media/131230/download. Accessed February 18, 2020.
19. Everaert K, Hervé F, Bosch R, et al. International Continence Society consensus on the diagnosis and treatment of nocturia. *Neurourol Urodyn.* 2019;38(2):478–498.
20. Oelke M, De Wachter S, Drake MJ, et al. A practical approach to the management of nocturia. *Int J Clin Pract.* 2017;71(11):e13027.
21. Weiss JP, Everaert K. Management of Nocturia and Nocturnal Polyuria. *Urology.* 2019;133S:24–33.
22. Kujubu DA. Nocturia in elderly persons and nocturnal polyuria. In *Online Curricula: Geriatric Nephrology*, Chapter 19. Washington DC: American Society of Nephrology; 2009.
23. Urology Care Foundation. Nocturia. 2020; https://www.urologyhealth.org/urology-a-z/n/nocturia. Accessed September 18, 2020.
24. Holm-Larsen T, Andersson F, van der Meulen E, Yankov V, Rosen RC, Nørgaard JP. The Nocturia Impact Diary: a self-reported impact measure to complement the voiding diary. *Value Health*. 2014;17(6):696–706.
25. ClinicalTrials.gov. Study investigating the impact burden of nocturia using the nocturia impact diary (IMPACT).ClinicalTrials.gov Identifier: NCT01552343. 2017; https://clinicaltrials.gov/ct2/show/results/NCT01552343. Accessed July 29, 2020.
26. ClinicalTrials.gov. Investigation of the superiority effect of desmopressin to placebo in terms of night voids reduction in nocturia adult female patients (COMFORT).ClinicalTrials.gov Identifier: NCT01223937. 2015; https://clinicaltrials.gov/ct2/show/results/NCT01223937. Accessed July 29, 2020.
27. Lose G, Lalos O, Freeman RM, van Kerrebroeck P; Nocturia Study Group. Efficacy of desmopressin (Minirin) in the treatment of nocturia: a double-blind placebo-controlled study in women. *Am J Obstet Gynecol.* 2003;189(4):1106–1113.
28. Dmochowski RR, Sanders SW, Appell RA, Nitti VW, Davila GW. Bladder-health diaries: an assessment of 3-day vs 7-day entries. *BJU Int.* 2005;96(7):1049–1054.
29. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 2nd ed. New York: Oxford University Press; 1995.
30. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973;33(3):613–619.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
32. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46.
33. Nunnally JC, Bernstein IH. *Psychometric Theory.* 3rd ed. New York: McGraw-Hill; 1994.
34. Williams V, McLeod L, Nelson L. Advances in the evaluation of longitudinal construct validity of clinical outcome assessments. *Ther Innov Regul Sci.* 2015;49(6):805–812.
35. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–159.

36. Luiz RR, Almeida RMVR. On the measurement of change in medical research. *Int J Stat Med Res.* 2012;1(2):144–147.
37. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res.* 2003; 12(4):349–362.
38. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(3):294–334.
39. 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(2):102–109.
40. Everaert K, Anderson P, Wood R, Andersson FL, Holm-Larsen T. Nocturia is more bothersome than daytime LUTS: Results from an Observational, Real-life Practice Database including 8659 European and American LUTS patients. *Int J Clin Pract.* 2018;72(6):e13091.

ACKNOWLEDGMENTS

Brian Samsell of RTI Health Solutions provided medical writing services. T. Roth and S. Ancoli provided advice on the qualitative research. Tine Kold Olesen of Ferring Pharmaceuticals A/S supported this project throughout the whole development phase.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication May 11, 2020
Submitted in final revised form November 9, 2020
Accepted for publication November 10, 2020
Address correspondence to: Valerie Williams, RTI Health Solutions, 3040 East Cornwallis Rd., Post Office Box 12194, Research Triangle Park, NC 27709-2194; Tel: (919) 316-3820; Fax: (919) 541-7222; Email: vwilliams@rti.org

DISCLOSURE STATEMENT

All authors have seen and approved this manuscript. Work for this study was performed at RTI Health Solutions. This study was funded by RTI Health Solutions, which received funding under a research contract with Ferring Pharmaceuticals A/S to conduct this study and provide editorial support in the form of manuscript writing, styling, and submission. Authors from Ferring Pharmaceuticals participated in designing the study, in analyzing and interpreting the data, in writing the manuscript, and in the decision to submit the article for publication. FLA and KVJ are employees of Ferring Pharmaceuticals A/S, a pharmaceutical research company active in nocturia. VW, SQ, CDR, SL, NW, and SY are employees of RTI Health Solutions, an independent nonprofit research organization, which received funding pursuant to a contract from Ferring Pharmaceuticals A/S to conduct the study.