

## SCIENTIFIC INVESTIGATIONS

# A New Single-Item Sleep Quality Scale: Results of Psychometric Evaluation in Patients With Chronic Primary Insomnia and Depression

Ellen Snyder, PhD; Bing Cai, PhD; Carla DeMuro, MS; Mary F. Morrison, MD, MS; William Ball, MD, PhD

Merck & Co., Inc., Kenilworth, New Jersey

**Study Objectives:** A single-item sleep quality scale (SQS) was developed as a simple and practical sleep quality assessment and psychometrically evaluated.

**Methods:** SQS measurement characteristics were evaluated using the Pittsburgh Sleep Quality Index (PSQI) and morning questionnaire-insomnia (MQI) according to prespecified analysis plans in separate clinical studies of patients with insomnia and depression. Patients with insomnia (n = 70) received 4 weeks' usual care with an FDA-approved hypnotic agent; patients with depression (n = 651) received 8 weeks' active or experimental therapy.

**Results:** Concurrent criterion validity (correlation with measures of a similar construct) was demonstrated by strong (inverse) correlations between the SQS and MQI (week 1 Pearson correlation  $-.76$ ) and PSQI (week 8 Goodman-Kruskal correlation  $-.92$ ) sleep quality items in populations with insomnia and depression, respectively. In patients with depression, stronger correlations between the SQS and PSQI core sleep quality components versus other items supported convergent/divergent construct validity (similarity/dissimilarity to related/unrelated measures). Known-groups validity was evidenced by decreasing mean SQS scores across those who sleep normally, those borderline to having sleep problems, and those with problems sleeping. Test-retest reliability (intraclass correlation coefficient) was  $.62$  during a 4-week period of sleep stability in patients with insomnia and  $.74$  in stable patients with depression (1 week). Effect sizes (standardized response means) for change from baseline were  $1.32$  (week 1) and  $.67$  (week 8) in populations with insomnia and depression, respectively. Mean SQS changes from baseline to week 8 convergently decreased across groups of patients with depression categorized by level of PSQI sleep quality improvement.

**Conclusions:** The SQS possesses favorable measurement characteristics relative to lengthier or more frequently administered sleep questionnaires in patients with insomnia and depression.

**Clinical Trial Registration:** Registry: ClinicalTrials.gov, Title: Treatment of Patients With Major Depressive Disorder With MK0869, Identifier: NCT00034983, URL: <https://clinicaltrials.gov/ct2/show/NCT00034983>

**Keywords:** depression, insomnia, instrument development, Morning Questionnaire, Pittsburgh Sleep Quality Index, psychometric evaluation, sleep, sleep quality scale

**Citation:** Snyder E, Cai B, DeMuro C, Morrison MF, Ball W. A new single-item sleep quality scale: results of psychometric evaluation in patients with chronic primary insomnia and depression. *J Clin Sleep Med*. 2018;14(11):1849–1857.

## BRIEF SUMMARY

**Current Knowledge/Study Rationale:** Currently available assessments of sleep quality are lengthy or require frequent administration, which may be burdensome for participants in non-sleep-related clinical trials. The single-item sleep quality scale was developed and validated to offer a practical alternative for sleep quality assessment in clinical settings.

**Study Impact:** The favorable measurement characteristics of the sleep quality scale demonstrated in this study (ie, strong concurrent criterion; convergent, divergent, and known-groups validity; effect size; and adequate test-retest reliability) support the reliability and validity of the sleep data derived from this sleep questionnaire and warrant its use as a sleep quality measure for assessment and treatment monitoring in clinical trials and clinical care.

## INTRODUCTION

Insomnia is a common health complaint characterized by difficulties in sleep initiation and/or maintenance, despite adequate sleep opportunities, which negatively affects daytime functioning.<sup>1</sup> United States epidemiological studies have reported prevalence estimates for insomnia ranging from 4% to 22%, depending on the diagnostic criteria used to define this condition.<sup>2</sup> Although it can arise in the absence of underlying conditions, insomnia also constitutes a core symptom of a multitude of other medical and psychiatric illnesses, such

as major depressive disorder, bipolar disorder, generalized anxiety disorders, posttraumatic stress disorder, and schizophrenia.<sup>3,4</sup> Problematic sleep is associated with adverse consequences on the patient's psychological, social, and cognitive functioning, which leads to a deterioration in the overall quality of life.<sup>5</sup> Given the significance of the burden posed by sleep disturbances on the affected individuals and its associated implications, tools for the measurement of sleep quality in clinical settings are of key importance to help determine whether a sleep complaint warrants further investigation and/or treatment and to assist in monitoring treatment.

Sleep can be assessed by measuring parameters such as sleep duration, sleep architecture, sleep latency, and the frequency and duration of awakenings throughout the night. The quantitative metrics may be measured using objective methods, including polysomnography and/or actigraphy,<sup>6,7</sup> or by way of sleep-rating questionnaires that capture the respondent's appraisal of these parameters. Sleep-rating questionnaires also capture ratings of the qualitative components of sleep quality, such as perceptions of sleep depth, rousing difficulties, and restfulness after sleep, in addition to other factors that could affect sleep quality, such as comorbid conditions and medication use.<sup>8</sup> The evaluation of the qualitative aspects of sleep experience is important, as sleep complaints can often persist despite normal values for quantitative measures of sleep.<sup>9</sup>

An evaluation of sleep quality is particularly important. In response to the need for evaluation instruments in insomnia research, a number of sleep questionnaires have been developed and validated.<sup>10</sup> However, despite their utility in the measurement of sleep quality, these tools are prone to several limitations when used in the context of clinical trials and may not always fulfill industry standards. For example, the Pittsburgh Sleep Quality Index (PSQI), a frequently used measure of sleep quality, was developed as a screening tool and may not be sensitive enough for detecting treatment differences in clinical trials.<sup>11</sup> In addition, the lengthy format of this questionnaire, which comprises 19 rating items, may preclude its administration to participants in clinical trials for practical reasons. A daily morning sleep diary questionnaire, hereafter referred to as the morning questionnaire-insomnia (MQI), is another multi-item tool for sleep quality evaluation requiring daily self-rating of sleep, which may be perceived as burdensome to patients. Both the PSQI and the MQI are limited by inclusion of only four possible response levels for the sleep quality question. This might limit the respondent's choice when grading his or her quality of sleep, undermining the accuracy of their answer, and may not allow the instrument to capture small or more subtle changes. As a result, despite the availability of other sleep quality questionnaires, there remains a need to develop and validate novel questionnaires that are specifically designed to facilitate sleep quality evaluation in the context of clinical trials.

The single-item sleep quality scale (SQS) is a sleep quality measure developed to provide a more pragmatic approach for the assessment of sleep quality in clinical settings, as compared with the commonly used standards of sleep quality evaluation. The SQS is a self-rated, global sleep quality assessment tool developed based on a literature review of key aspects of sleep quality, critical components of the PSQI and the MQI, and direct expert and patient input. The single-item format enables a patient-reported rating of sleep quality over a 7-day recall period without greatly increasing the patient's burden. The use of a discretizing visual analog scale (VAS) increases the potential for a more sensitive measurement. Based on rigorous development, the scale appears to be both face and content valid. The aim of this study was to establish the validity, reliability, responsiveness, and potential to detect clinically significant changes of the SQS, relative to the MQI and PSQI, in two clinical populations with sleep impairment;

specifically, patients with chronic insomnia and patients with major depressive disorder.

## METHODS

### Data Sources

The SQS was validated based on two clinical studies (an insomnia study and a depression study), which were conducted in accordance with principles of Good Clinical Practice and approved by the appropriate institutional review boards and regulatory agencies. All participants provided informed consent.

The insomnia study data (**Table 1**) were obtained from a 4-week, randomized, parallel-group, multicenter study of patient-reported sleep effects following usual care with any United States Food and Drug Administration (FDA)-approved hypnotic agent following a 1-week washout period (Protocol EP4003-006 and EP4003-009, referred to as Study 006 and Study 009, respectively). Study 006 was designed to evaluate measurement characteristics of an electronic device for collection of diary and questionnaire data in patients with primary insomnia, along with efficiency, costs, and quality associated with electronic collection. Eligible outpatients ( $n = 70$ ) were aged 30 to 75 years and were receiving an FDA-approved prescription hypnotic agent as a part of usual care for chronic primary insomnia diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria. In addition, patient-reported total sleep time was  $\leq 6$  hours at the first visit and median total sleep time was  $\leq 6$  hours on the first 7 nights of the washout period. Patients were randomized equally into two study arms and stratified by age (younger than 60 years, 60 years or older) and education (0 to 12 years, more than 12 years). Study arm 1 comprised at-home and on-site patients using small-format electronic devices. Study arm 2 included at-home and on-site patients using paper questionnaires. The MQI was administered daily throughout the study.

Evaluations using the SQS were undertaken at screening, at the end of the 1-week washout period, and at the end of weeks 1 and 4 of the usual-care treatment period as part of Study 009, a substudy of Study 006. Study 009 aimed to examine burden of illness as an exploratory endpoint, evaluating health resource utilization, patient-reported work productivity, functional outcomes, health status, and disability. Although all patients needed to agree to participate in both the parent study and substudy, the protocols were separated for administrative convenience to facilitate collection of resource and data-quality metrics for Study 006, without being affected by the ancillary data collected in Study 009.

Data for patients with major depressive disorder ( $n = 651$ ) (**Table 1**) were obtained from the initial 8-week treatment period of a randomized, double-blind, parallel-group, 12-month international study evaluating the long-term safety and tolerability of ascending doses of the substance P antagonist aprepitant (80, 160, and 240 mg) versus the active comparator paroxetine hydrochloride (20 mg) (NCT00034983; Protocol MK0869-066; referred to as Study 066). The study was conducted between October 2001 and December 2003. Eligible

**Table 1**—Patient baseline characteristics.

	Insomnia Study (n = 70)	Depression Study (n = 651)
Sex, n (%)		
Female	40 (57)	358 (59)
Male	30 (43)	266 (41)
Race, n (%)		
Black	14 (20)	30 (5)
White	53 (76)	536 (82)
Other	3 (4)	85 (13)
Age (years), mean (SD)	50.2 (11.8)	45.1 (16.2)
PSQI total sleep time (hours), mean (SD)	4.6 (.9)	5.8 (1.7)
PSQI sleep latency (minutes), mean (SD)	55.3 (44.0)	42.6 (41.8)

PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation.

patients were aged 18 years or older (17% of patients were 65 years of age or older) and had a diagnosis of major depressive disorder based on DSM-IV criteria. In addition, to be included in the study, patients were required to demonstrate a score of 18 or higher on the total of the first 17 items of the 21-item Hamilton Depression Rating Scale at both the prestudy and baseline visits. In addition to the primary and secondary endpoints that assessed safety and efficacy of aprepitant, the assessment of the concurrent validity of the SQS, as compared with the PSQI, was prespecified as an additional endpoint in Study 066. Patients completed the SQS and the PSQI at the baseline visit and at the end of weeks 1 and 8. If a patient discontinued the study prior to the end of the week 8 visit, the patient completed the SQS and PSQI at the final study visit. The SQS was administered at sites in the United States (~60% of patients).

## Description of the Sleep Scales

### The SQS

The SQS is a self-administered questionnaire that incorporates a discretizing VAS. The questionnaire instructions direct the respondent to rate the overall quality of sleep over a 7-day recall period on a discretizing VAS, whereby the respondent marks an integer score from 0 to 10, according to the following five categories: 0 = terrible, 1–3 = poor, 4–6 = fair, 7–9 = good, and 10 = excellent (**Figure 1**). When rating their sleep quality, respondents are instructed to consider the following core components of sleep quality: how many hours of sleep they had, how easily they fell asleep, how often they woke up during the night (except to go to the bathroom), how often they woke up earlier than they had to in the morning, and how refreshing their sleep was.

### The PSQI

The originally published PSQI assesses sleep quality during the previous month and comprises 19 individual items used to derive the following seven component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency,

**Figure 1**—Sleep quality scale instructions.

**INSTRUCTIONS:**

- The following question refers to your overall sleep quality for the **majority** of nights in the **past 7 days ONLY**.
- Please think about the quality of your sleep **overall**, such as how many hours of sleep you got, how easily you fell asleep, how often you woke up during the night (except to go to the bathroom), how often you woke up earlier than you had to in the morning, and how refreshing your sleep was.

1. During the **past 7 days**, how would you rate your sleep quality overall?  
(Please mark only 1 box)

Terrible      Poor      Fair      Good      Excellent

0    1    2    3    4    5    6    7    8    9    10

sleep disturbances, use of sleeping medication, and daytime dysfunction.<sup>11</sup> Each component's score can take the value of 0, 1, 2, or 3. The sum of the scores of the seven components is used to compute the global sleep quality score, ranging in value from 0 to 21. A score of 0 indicates no sleep difficulties, while more-severe sleep difficulties correspond to higher values for the global and component scores. The PSQI used in the current study is a validated, modified version of the original PSQI, evaluating sleep quality over a 1-week recall period instead of a 4-week recall period.

### The MQI

The MQI is a 12-item sleep diary questionnaire that is completed by patients daily. It includes patient reports of time at which they got into and out of bed (items 1 and 2), whether they fell asleep (item 3), and if so, the sleep latency (item 4), number and duration of awakenings (items 5 and 6), and total sleep time (item 7). Based on information provided by the respondent, a sleep efficiency value (ie, the ratio of total sleep time over the number of hours spent in bed) can be derived. Sleep quality (item 8) and ability to concentrate in the morning (item 12) are each rated on a 4-point scale with the following levels: poor (4), fair (3), good (2), and excellent (1). Ease of sleep (item 9) and the feeling of refreshment in the morning (item 11) are self-rated using a VAS from 0 to 100 mm (0 = very, 100 = not at all). Patients also report any unusual events that disturbed sleep (yes or no, with a description of the event if yes; item 10).

## Validation Analysis Methods

Treatment groups were combined within each study for the purposes of the validation analyses. Descriptive statistics were computed for baseline patient demographic characteristics and for the SQS, MQI, and PSQI items (components and global score) at each timepoint. For continuous variables, mean, standard deviation (SD), 25th percentile, median, 75th percentile, minimum, and maximum were computed. For categorical variables, frequencies were calculated. The SQS data were summarized with both continuous and categorical formats (frequency distributions for both the 11 response options and the 5-level categories).

In the insomnia study, observed case data were used because few observations were missing. Because the dropout

**Table 2**—Concurrent criterion validity.

Population	Statistic	Timepoint	n	Estimate
Insomnia study	Pearson correlation between SQS and average MQI sleep quality (item 8)	Week 1	66	-.76
		Baseline	639	-.87
Depression study	Goodman-Kruskal correlation between SQS and PSQI sleep quality (item 6)	Week 1	634	-.88
		Week 8	648	-.92

MQI = morning questionnaire-insomnia, PSQI = Pittsburgh Sleep Quality Index, SQS = sleep quality scale.

rate by week 8 in the depression study was 40%, a modified last-observation-carried-forward (MLOCF) approach was used to impute missing data at week 8: if the SQS or all items on the PSQI were missing at week 8, these items were imputed by carrying forward, to week 8, the patient's previous concurrent pair of SQS and PSQI questionnaires occurring in the treatment phase (ie, week 1 or early discontinuation visit). However, the reliability analysis was based on observed case data.

### Measurement Characteristics

Measurement characteristics of the SQS were assessed relative to established measures of sleep quality, including the MQI and PSQI. The PSQI included a 1-week recall. Analyses were conducted according to a prespecified validation analysis plan.

### Concurrent Criterion Validity

Concurrent criterion validity demonstrates the extent to which an instrument correlates a measure assessing a similar construct. The common sleep criterion considered was sleep quality (questionnaire items 8 and 6 on the MQI and PSQI, respectively). In the insomnia study, concurrent criterion validity had a cross-section evaluation by deriving a Pearson correlation between the SQS score and the 7-day average of the MQI item 8 at weeks 1 and 4. In the depression study, concurrent criterion validity was assessed by calculating a Goodman-Kruskal correlation coefficient for the SQS and the PSQI sleep quality (item 6) at baseline, week 1, and week 8.<sup>12</sup> Correlations  $\geq .4$  are considered to represent adequate concurrent criterion validity.

### Convergent/Divergent Construct Validity

Convergent construct validity is the extent to which a measure correlates with another measure that it should theoretically be related to, whereas divergent construct validity is the extent to which a measure is relatively dissimilar from another that it is not expected to be related to. In the case of the SQS, convergent construct validity was examined by correlating the SQS score with core components of the sleep quality construct in the PSQI, namely sleep time, latency, efficiency, quality, awakenings, ease falling asleep, and degree of feeling refreshed. Divergent construct validity was assessed by correlating the SQS score with the other sleep items of the PSQI, which are not expected to be significantly related to sleep quality. Correlations between SQS score and the PSQI items, components, and total scores were assessed at baseline, week 1, and week 8 in the depression study. For continuous variables (eg, total sleep

time, sleep latency, and global score), a Pearson correlation coefficient was computed. For all other PSQI sleep items, a Goodman-Kruskal correlation was determined.

### Known-Groups Construct Validity

Known-groups construct validity examines the ability of a measure to discriminate between different groups of individuals. Normal and problematic sleep can be differentiated based on the PSQI global scores. Individuals with scores less than 5 are considered to sleep normally, whereas those with scores higher than 8 are deemed to have sleep problems.<sup>13</sup> In the depression study, the SQS score (95% confidence interval [CI]) was computed for those who sleep normally (PSQI global score less than 5 points), those borderline to having sleep problems (PSQI global score 5 to 8 points), and those with sleep problems (PSQI global score higher than 8 points) at baseline, week 1, and week 8. A test for trend in SQS scores across the different groups at each timepoint was also performed.

### Test-Retest Reliability

Test-retest reliability reflects the temporal stability of a measure when no change is anticipated. In view of the lack of a placebo group in both the insomnia and depression studies, two approaches were considered to derive reliability metrics, and these two approaches differ in the way stable patients are defined.

First, test-retest reliability was estimated by calculating intraclass correlation coefficients (ICCs) for all patients during an anticipated period of stability.<sup>14</sup> Because most improvement in sleep was anticipated during the first week of treatment, the most stable period with regard to sleep would be between week 1 and week 4 in the insomnia study and between week 1 and week 8 in the depression study.

Alternatively, ICCs were calculated for a subset of patients deemed to be stable because they did not report change on a validated sleep measure. In the insomnia study, stable patients were defined as those demonstrating a change of within  $\pm 1$  on the MQI sleep quality item 8. Test-retest reliability for stable patients with insomnia was computed for baseline versus week 1 and for week 1 versus week 4. In the depression study, stability was defined as no change in sleep quality as evaluated based on the PSQI sleep quality item 6. Test-retest reliability in stable patients with depression was assessed for baseline versus week 1 and for week 1 versus week 8. ICCs of .5 to .7 are considered acceptable, whereas ICCs  $\geq .7$  are considered good.

**Table 3**—Convergent/divergent construct validity: correlations between SQS and PSQI items at week 8 in patients with depression.

PSQI Item	Correlation Coefficient*	Standard Deviation	95% Confidence Interval
Sleep latency (item 2) †	-.38	.04	-.46, -.30
Total sleep time (item 4) †	.49	.06	.37, .61
Cannot fall asleep ≤ 30 minutes (item 5a) †	-.51	.03	-.58, -.45
Disturbance items			
Awakenings (item 5b) †	-.56	.03	-.62, -.50
Use bathroom (item 5c)	-.20	.04	-.28, -.13
Cannot breathe (item 5d)	-.29	.06	-.40, -.17
Cough/snore (item 5e)	-.17	.06	-.28, -.05
Feel too cold (item 5f)	-.25	.05	-.34, -.15
Feel too hot (item 5g)	-.23	.04	-.31, -.14
Had bad dreams (item 5h)	-.23	.05	-.32, -.14
Have pain (item 5i)	-.23	.05	-.33, -.14
Other (item 5j)	-.32	.05	-.42, -.23
Sleep quality (item 6)	-.92	.01	-.95, -.89
Sleep medication (item 7)	-.32	.06	-.43, -.21
Trouble staying awake (item 8)	-.21	.04	-.29, -.12
Enthusiasm (item 9)	-.44	.04	-.51, -.37
Component scores			
Sleep quality (component 1)	-.92	.01	-.95, -.89
Sleep latency (component 2)	-.48	.03	-.55, -.41
Sleep duration (component 3)	-.55	.03	-.62, -.49
Habitual sleep efficiency (component 4)	-.48	.04	-.55, -.40
Sleep disturbances (component 5)	-.53	.04	-.61, -.45
Sleep medications (component 6)	-.32	.06	-.43, -.21
Daytime dysfunction (component 7)	-.40	.04	-.48, -.32
Global score	-.72	.02	-.76, -.68

\* = Pearson correlation coefficient for continuous items (ie, total sleep time, sleep latency, and global score); Goodman-Kruskal correlation coefficients for all other components of sleep quality. † = core component of sleep quality (ie, aspect of sleep that patients were asked to consider when responding to the SQS questions). PSQI = Pittsburgh Sleep Quality Index, SQS = sleep quality scale.

### Effect Size

The effect size reflects the ability of a measure to detect a change (eg, in response to treatment). For the SQS, effect size was estimated using the following two approaches<sup>14</sup>: first, by calculating a responsiveness statistic, which is the mean change from baseline in sleep quality for all patients divided by the SD for change from baseline for stable patients; second, by computing the standardized response mean, defined as the mean change from baseline for all patients divided by the SD for change from baseline for all patients. In the insomnia study, the standardized response means for change from baseline to treatment week 1 and week 4 were computed. The effect size in the depression study was calculated as the responsiveness statistic or the standardized response mean from baseline to week 1 and week 8. Effect sizes of .2, .5, and .8 are considered small, moderate, and large, respectively.

### Detection of Clinically Meaningful Differences

The ability of the SQS to detect clinically meaningful changes in sleep quality was investigated. The observed distributions

of changes in the SQS were summarized in defined groups based on change from baseline in PSQI sleep quality (item 6) as follows: greatly improved (−2 or −3), somewhat improved (−1 change), no change (0 change), somewhat worsened (+1 change), greatly worsened (+2 or +3 change). The observed distributions of changes in the SQS were also summarized for each cell in a cross-classification table of PSQI sleep quality at baseline versus weeks 1 and 8 in the depression study. The following statistics were computed for each of the groups defined above: n, mean, SD, minimum, median, and maximum.

## RESULTS

### Evaluation of Measurement Characteristics of the SQS

#### Concurrent Criterion Validity

Overall, there was excellent concurrent validity of the SQS with the MQI and PSQI as evidenced by the strength of the correlation coefficients derived from both the insomnia and

**Table 4**—Known-groups validity assessed by SQS scores for normal, borderline, and problem sleepers among patients with depression.

Timepoint	Sleep Category*	n	Mean	SD	95% CI
Baseline	Normal	53	6.8	1.7	6.3, 7.3
	Borderline	163	5.2	1.7	5.0, 5.5
	Problem	361	3.1	1.6	3.0, 3.3
Week 1	Normal	146	7.3	1.6	7.0, 7.5
	Borderline	221	5.5	1.8	5.3, 5.7
	Problem	222	3.7	1.9	3.4, 3.9
Week 8	Normal	202	7.5	1.6	7.2, 7.7
	Borderline	211	5.9	1.7	5.6, 6.1
	Problem	193	3.8	2.0	3.6, 4.1

\* = categories defined by PSQI global scores: normal, borderline, and problem sleepers are defined as those with scores less than 5, 5 to 8, and higher than 8, respectively. Test for trend  $P \leq .001$  for all timepoints. CI = confidence interval, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, SQS = sleep quality scale.

**Table 5**—Test-retest reliability.

Population	Test-Retest Reliability	n	Estimate
Insomnia study	Intraclass correlation for week 1 versus week 4 (all patients)*	70	.62
Depression study	Intraclass correlation for baseline versus week 1 (stable patients) †	296	.74
	Intraclass correlation for week 1 versus week 8 (stable patients) †	213	.55
	Intraclass correlation for week 1 versus week 8 (all patients)*	444	.30

\* = all patients during a stable period of treatment. † = because there was no stable period of treatment for patients with depression, the intraclass correlation coefficient was computed for a subset of patients defined as stable based on that they did not demonstrate a change from baseline on PSQI sleep quality item 6 in the considered period. PSQI = Pittsburgh Sleep Quality Index.

the depression study populations (**Table 2**). In patients with insomnia, the Pearson correlation for the SQS relative to the average MQI sleep quality (item 8) at week 1 was  $-.76$ . In patients with depression, Goodman-Kruskal correlation of SQS relative to PSQI sleep quality (item 6) was evaluated at baseline, week 1, and week 8 and correlation coefficients were  $-.87$ ,  $-.88$ , and  $-.92$ , respectively.

### Convergent/Divergent Construct Validity

Convergent and divergent construct validity were investigated by correlating the SQS score with different items of the PSQI at weeks 1 and 8 in the depression study. **Table 3** reports the correlation results for week 8. The Pearson correlation between the SQS and the global PSQI score, which takes into account all items of the PSQI, was strong (correlation coefficient:  $-.72$ ). The results of construct validity analyses performed using week 1 data were generally consistent with the findings at week 8 (data not shown).

Convergent construct validity evaluates the correlation between the SQS and PSQI scores with respect to the core components of sleep quality (ie, those aspects of sleep that patients were asked to consider when responding to the SQS question). Correlation coefficients for the core components of sleep quality at week 8 were  $-.38$  for sleep latency,  $.49$  for total sleep time,  $-.51$  for difficulty falling asleep within 30 minutes, and  $-.56$  for frequency of awakening (**Table 3**).

Divergent construct validity measures the correlation between unrelated sleep items in the two questionnaires.

Correlations between the SQS and PSQI were weaker for PSQI items not related to core sleep quality, with correlation coefficients ranging from  $-.17$  to  $-.44$ , compared with core components of sleep quality.

Given that higher scores on the PSQI indicate more severe sleep issues (with the exception of total sleep time), whereas higher SQS scores denote better sleep, the negative correlations between the PSQI and the SQS scores (except total sleep time) were expected.

### Known-Groups Construct Validity

The SQS scores for patients with depression, categorized as those who sleep normally, those borderline to having sleep problems, and those with sleep problems based on PSQI global scores of less than 5, 5 to 8, and higher than 8, respectively,<sup>13</sup> were compared at baseline, week 1, and week 8 (**Table 4**). At each timepoint, the SQS score was highest for those who sleep normally (6.8 to 7.5 across timepoints), intermediate for those borderline to having sleep problems (5.2 to 5.9), and lowest for those with sleep problems (3.1 to 3.8). At all timepoints, the differences in mean SQS scores between groups were significant ( $P < .001$ ) and the 95% CI for the scores of the different groups did not overlap. Based on these results, the SQS is a valid measure when used to discriminate between different sleep categories.

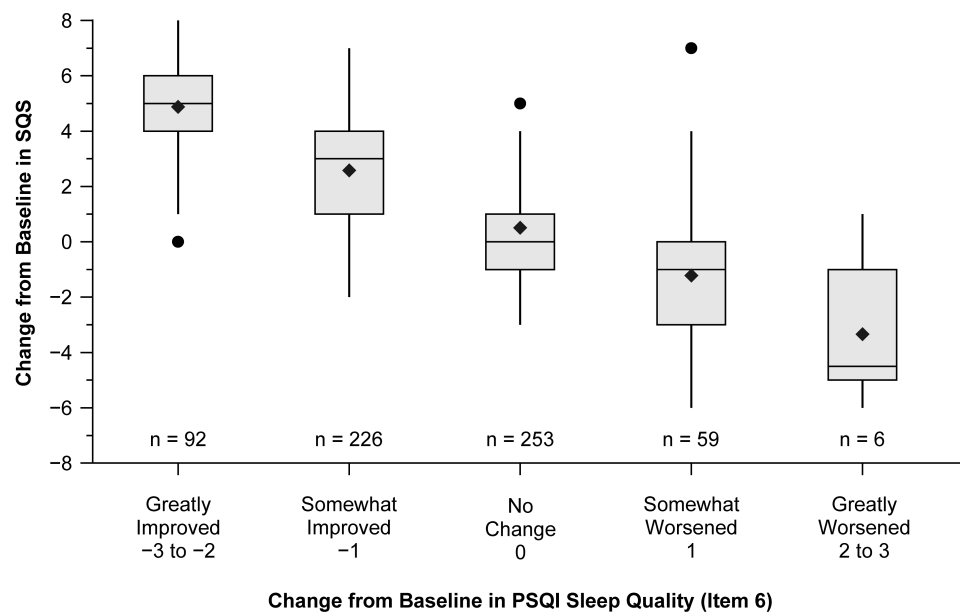
### Test-Retest Reliability

Test-retest reliability of the SQS was evaluated by repeated measurements during a period of sleep stability (**Table 5**). In

**Table 6**—Effect size.

Population	Effect Size	n (All Patients)	n (Stable Patients)	Estimate
Insomnia study	Standardized response mean for change from baseline to week 1	68	—	1.32
	Standardized response mean for change from baseline to week 4	62	—	1.75
Depression study	Standardized response mean for change from baseline to week 1	630	—	.51
	Standardized response mean for change from baseline to week 8*	643	—	.67
	Responsiveness statistic for change from baseline to week 1	630	294	.72
	Responsiveness statistic for change from baseline to week 8*	643	253	1.10

Standardized response mean is defined as the mean change from baseline for all patients divided by the SD for change from baseline for all patients. The responsiveness statistic is defined as the mean change from baseline for all patients divided by the SD for change from baseline for stable patients (no change from baseline based on PSQI item 6). \* = week 8 modified last-observation-carried-forward. PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation.

**Figure 2**—Anchor-based analysis of clinically meaningful differences at week 8 in patients with depression.

Boxes represent interquartile ranges. Vertical lines represent the minimum data point within the 25th percentile minus 1.5 times the interquartile range and maximum data point within the 75th percentile plus 1.5 times the interquartile range. Horizontal lines within the boxes represent the medians. Diamonds represent means. Dots represent outliers. PSQI = Pittsburgh Sleep Quality Index, SQS = sleep quality scale.

patients with insomnia, the ICC for week 1 versus week 4 (considered as a stable period of sleep) for all patients ( $n = 70$ ) was moderate and estimated at .62.

A similar calculation was not considered an appropriate measure of reliability in the depression study because the 7-week period following week 1 was not a stable sleep period with regard to sleep quality for all patients (ie, sleep quality improved beyond the first week of treatment). Among subgroups of patients with depression who demonstrated stability in sleep quality based on the PSQI sleep quality assessment (item 6), the 1-week reliability (baseline to week 1;  $n = 296$  stable patients) was strong, with an ICC of .74, and the 7-week reliability (week 1 to week 8;  $n = 213$  stable patients) was also acceptable (ICC = .55).

### Responsiveness

The ability of the SQS to detect changes in sleep quality in response to treatment was evaluated in both the insomnia and

depression studies. The value of the standardized response means for change from baseline to week 1 and week 4 in patients with insomnia were 1.32 and 1.75, respectively, which denotes a strong effect size (Table 6). In the depression study, the standardized response means from baseline to week 1 and from baseline to week 8 were moderate with respective values of .51 and .67. The responsiveness statistic was higher than the standardized response mean, indicating a moderate effect size at week 1 (.72) and a large effect size at week 8 (1.1; MLOCF) (Table 6).

### Detection of Clinically Meaningful Change

The ability of the SQS to measure clinically meaningful differences in sleep quality was investigated in the depression study. Patients with depression were assigned to different categories based on the level of improvement in the sleep quality component (item 6) of the PSQI from baseline, as follows: greatly improved (change of  $-2$  to  $-3$ ), somewhat improved

(change of -1), no change (change of 0), somewhat worsened (change of +1), and greatly worsened (change of +2 to +3). At week 8, the mean changes from baseline in SQS scores were highest (mean SQS change of 4.9) among patients with greatly improved sleep quality on the PSQI (item 6) and decreased across groups with decreasing PSQI (item 6) improvement level (mean SQS change of 2.6, .5, -1.2, and -3.3 in the somewhat improved, no change, somewhat worsened, and greatly worsened categories, respectively) (**Figure 2**). The data for the change of SQS scores from baseline to week 1 showed a similar trend. These results demonstrate the utility of the SQS in discerning clinically meaningful differences in sleep quality among groups.

## DISCUSSION

In the current study, a rigorous psychometric evaluation was implemented to assess the measurement properties of the SQS relative to other measures of sleep, namely the MQI and PSQI. The findings indicate that the SQS possesses excellent concurrent criterion validity. There were strong correlations between the SQS and the sleep quality items of the MQI and PSQI in patients with insomnia and patients with depression, respectively. The direction of the association was negative because better sleep quality is associated with a lower score on the MQI and PSQI, but a higher score on the SQS. Correlations between the SQS and core components of sleep quality of the PSQI were much stronger compared with other items of the scale, supporting the convergent and divergent construct validity of the SQS. The strong correlation of the SQS with the PSQI global score, which takes into account all items of the PSQI, also supports the construct validity of the SQS.

Known-groups validity was corroborated in the depression study by the clear significant differences ( $P < .001$ ) in mean SQS scores among patients classified as sleeping normally, borderline to having sleep problems, and having sleep problems based on PSQI scores. The SQS scores decreased across groups with increasing severity of sleep impairment.

The SQS demonstrated moderate to strong intraclass correlations during periods of relative sleep stability; however, neither study design was optimal for assessing test-retest reliability because a placebo group was not included as a study arm and the SQS was not administered on 2 consecutive weeks during the treatment period of either study to give repeated measurements during a short, stable period. Because improvements in sleep quality beyond the first week of treatment were observed in the depression study, 7-week test-retest reliability among all patients was not considered a valid measure of SQS reliability, and, hence, test-retest reliability was investigated in a subset of patients demonstrating relative sleep stability during this period. The ICCs for 1-week and 7-week test-retest reliability in patients with depression who demonstrated sleep stability based on the PSQI quality item may overestimate the true reliability of the SQS, because less change may occur among stable patients than might be expected with patients in a placebo group.

The ability of the SQS to detect change over time in response to treatment was evidenced by the strong effect size in the population with insomnia (baseline to week 1) and the moderate effect sizes observed at weeks 1 and 8 in the population with depression, when the effect size was expressed as the standardized response mean.<sup>15</sup> As anticipated, the standardized response means were smaller than the corresponding responsiveness statistics, given that the denominator reflects variation due to both treatment and chance.

Observed changes in mean SQS scores from baseline decreased across groups of patients who greatly improved, moderately improved, were unchanged, moderately worsened, and greatly worsened on PSQI sleep quality (item 6). These data provide benchmarks for clinically meaningful changes in the SQS scores.

This study has several limitations. As discussed previously, the designs of the clinical trials used as data sources for the SQS validation analysis were not optimal for the assessment of reliability because they did not include a placebo arm where participants would be anticipated to have a period of stability with respect to sleep. However, sleep often improves in patients with insomnia receiving placebo in clinical trials, so a placebo arm may not address the problem with sleep stability.<sup>16</sup> In addition, the current validation did not address content validity or sensitivity. Finally, the size of the insomnia study population ( $n = 70$ ) was relatively small.

In conclusion, this psychometric evaluation provides evidence that the single-item SQS possesses favorable measurement characteristics to assess sleep quality as demonstrated by the strong concurrent criterion, convergent/divergent validity, known-groups validity, acceptable reliability, and ability to measure responsiveness and clinically meaningful changes in sleep quality relative to the more frequently administered MQI or the lengthier PSQI. The findings support the utility of the SQS as a practical sleep measure that can effectively gauge sleep quality without significantly increasing the burden of clinical trial participants. It is also possible that the SQS may be useful in the clinical setting, for example, to identify a problem that might then warrant further evaluation using a more comprehensive assessment such as the PSQI. Further evaluation of this tool in other populations with sleep disturbance such as those with chronic pain, as well as in the general population, may be useful.

## ABBREVIATIONS

CI, confidence interval  
 DSM, Diagnostic and Statistical Manual of Mental Disorders  
 FDA, Food and Drug Administration  
 ICC, intraclass correlation coefficient  
 MLOCF, modified last-observation-carried-forward  
 MQI, morning questionnaire-insomnia  
 PSQI, Pittsburgh Sleep Quality Index  
 SD, standard deviation  
 SQS, sleep quality scale  
 VAS, visual analog scale



## REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
2. Roth T, Coulouvrat C, Hajak G, et al. Prevalence and perceived health associated with insomnia based on DSM-IV-TR; International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; and Research Diagnostic Criteria/International Classification of Sleep Disorders, Second Edition criteria: results from the America Insomnia Survey. *Biol Psychiatry*. 2011;69(6):592–600.
3. Krystal AD. Psychiatric disorders and sleep. *Neurol Clin*. 2012;30(4):1389–1413.
4. Doghramji K. The epidemiology and diagnosis of insomnia. *Am J Manag Care*. 2006;12(8 Suppl):S214–S220.
5. Szentkiralyi A, Madarasz CZ, Novak M. Sleep disorders: impact on daytime functioning and quality of life. *Expert Rev Pharmacoecon Outcomes Res*. 2009;9(1):49–64.
6. Littner M, Kushida CA, Anderson WM, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*. 2003;26(3):337–341.
7. Littner M, Hirshkowitz M, Kramer M, et al. Practice parameters for using polysomnography to evaluate insomnia: an update. *Sleep*. 2003;26(6):754–760.
8. Zhang L, Zhao ZX. Objective and subjective measures for sleep disorders. *Neurosci Bull*. 2007;23(4):236–240.
9. Krystal AD, Edinger JD. Measuring sleep quality. *Sleep Med*. 2008;9 Suppl 1:S10–S17.
10. Lomeli HA, Perez-Olmos I, Talero-Gutierrez C, et al. Sleep evaluation scales and questionnaires: a review. *Actas Esp Psiquiatr*. 2008;36(1):50–59.
11. Buysse DJ, Reynolds CF, 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.
12. Argenti A. *Analysis of Ordinal and Categorical Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2010.
13. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. *J Psychosom Res*. 1998;45(1):5–13.
14. Fayes PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. England: John Wiley & Sons, Ltd.; 2000.
15. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
16. Huedo-Medina TB, Kirsch I, Middlemass J, Klonizakis M, Siriwardena AN. Effectiveness of non-benzodiazepine hypnotics in treatment of adult insomnia: meta-analysis of data submitted to the Food and Drug Administration. *BMJ*. 2012;345:e8343.

## ACKNOWLEDGMENTS

Medical writing support, under the direction of the authors, was provided by Hicham Naimy, PhD, of CMC Affinity, a division of Complete Medical Communications, Inc., Hackensack, NJ, USA, funded by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, in accordance with Good Publication Practice (GPP3) guidelines. Christopher Lines, PhD from Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA provided comments on a draft. Dr. Morrison's current affiliation is: Lewis Katz School of Medicine at Temple University, Philadelphia, PA, USA.

## SUBMISSION &amp; CORRESPONDENCE INFORMATION

**Submitted for publication March 6, 2018**

**Submitted in final revised form May 16, 2018**

**Accepted for publication June 5, 2018**

Address correspondence to: Ellen Snyder, UG1C-46, 351 N. Sumneytown Pike, North Wales, PA 19454; Tel: (267) 305-7936; Fax: (267) 305-6395; Email: ellen\_snyder@merck.com

## DISCLOSURE STATEMENT

Funding for this research was provided by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA. All authors are current or former employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA. ES owns stock in Merck & Co., Inc., Kenilworth, NJ, USA. Dr. Morrison is a consultant to Marinus Pharmaceuticals, Radnor, PA, USA (no direct conflict with this manuscript). All authors have reviewed and approved the content of this manuscript. Work for this study was performed at Merck & Co., Inc., Kenilworth, NJ, USA.