ELSEVIER

Sleep Medicine 3 (2002) 99–108

SLEEP MEDICINE

www.elsevier.com/locate/sleep

Original article

A new questionnaire to detect sleep disorders ☆

Thomas Roth[a,*], Gary Zammit[b], Clete Kushida[c], Karl Doghramji[d], Susan D. Mathias[e], Josephine M. Wong[f], Daniel J. Buysse[g]

[a]*Henry Ford Sleep Center, 2799 West Grand Boulevard, CFP-3 Detroit, MI 48202, USA*
[b]*St. Luke's-Roosevelt Hospital Center, New York, NY, USA*
[c]*Stanford Sleep Disorders Clinic & Research Center, Stanford, CA, USA*
[d]*Thomas Jefferson University, Philadelphia, PA, USA*
[e]*The Lewin Group, San Francisco, CA, USA*
[f]*Pharmacia Corporation, Skokie, IL, USA*
[g]*University of Pittsburgh School of Medicine, Pittsburgh, PA, USA*

Received 3 January 2001; received in revised form 20 March 2001; accepted 22 March 2001

Abstract

Objectives: Sleep disorders remain largely undiagnosed in the general population. The current study assess7ed whether the Global Sleep Assessment Questionnaire (GSAQ) could: (1), distinguish between sleep disorders (including no sleep disorder); (2), be a reliable and valid sleep disorder screener; and (3), serve as a practical, user-friendly screening tool for primary care and sleep centers.

Methods: Two hundred and twelve adults from five sleep centers and two primary care clinics completed the GSAQ and received confirmed diagnoses from a sleep specialist. Of the 212 patients, 139 (65.6%) had at least one sleep disorder, 60 (28.3%) had two or more sleep disorders, and 13 (6.1%) had no confirmed sleep disorder. Ninety-one (43%) individuals completed the GSAQ a second time for reliability testing. Scores for each sleep disorder including, but not limited to, primary insomnia (I), insomnia associated with a mental disorder (IME), obstructive sleep apnea (OSA), periodic limb movement (PLM), and parasomnia (P) were computed. The sensitivity and specificity were estimated using comprehensive clinical diagnosis as the gold standard and mean domain scores as a cutpoint.

Results: The mean participant age was 45 years, 52% were female. Observed frequencies were: 36 (I), 14 (IME), 31 (OSA), 7 (PLM) and 4% (P). Test–retest reliability ranged from 0.51 to 0.92. Pearson correlation coefficients suggested that the GSAQ discriminated between diagnoses. The sensitivities and specificities were 79/57, 83/51, 93/58, 93/52, and 100/49 for I, IME, OSA, PLM, and P, respectively.

Conclusions: Our findings suggest that the GSAQ can aid in recognizing sleep disorders. Future studies should focus on characterizing its predictive values in primary care settings.

Keywords: Sleep disorders; Global sleep assessment questionnaire; Predictive value

1. Introduction

Sleep disorders have an estimated prevalence of 15–27% in the adult population [1,2]. According to the National Commission on Sleep Disorders Research, approximately 70 million Americans suffer from sleep-related problems. Of those, 40 million have chronic sleep disorders [1]. Obstructive sleep apnea (OSA) and disorders associated with insomnia account for the majority of sleep disorders, occurring in as many as 9% of women, 24% of men [3], and 30–36% of the population, respectively [4].

In addition to their high prevalence rates, sleep problems are associated with health problems, functioning and well-being, work-related indicators and health care expenditures [5,6]. A number of recent studies have demonstrated that individuals with current sleep problems report significantly poorer health, less energy and worse cognitive functioning than those categorized as having no sleep problem [5,7]. In another study, individuals with insomnia resulting in daytime dysfunction reported significantly lower quality-of-life scores on each of the individual eight domains and summary scores of the Short Form-36 (SF-36), as well as greater resource utilization due to more frequent ER visits, physician visits and use of prescription and/or over-the-counter drugs [6].

Despite these consequences, sleep disorders are not widely detected or treated [5]. Possible reasons for this finding may be the lack of training in the recognition of sleep disorders, the uncertainty of how to treat and/or refer patients with this condition, or because of the failure of patients and/or providers to discuss sleep problems during a health care visit.

☆ This study was supported by Pharmacia Corporation.

* Corresponding author. Tel.: +1-313-664-3571; fax: +1-313-664-3567.
E-mail address: troth1@hfhs.org (T. Roth).

1389-9457/02/$ - see front matter
PII: S1389-9457(01)00131-9

Currently, a limited number of screening tools are available to detect some sleep disorders in adults [8–14]. The Berlin questionnaire [8], which was designed to identify patients as being at 'high' or 'low' risk for OSA, assesses the patient's risk level based on approximately 11 questions addressing three symptom categories: snoring, sleepiness, and high blood pressure/weight. The Sleep Disorders Questionnaire (SDQ) [9] was developed to diagnose four categories of sleep disorders: OSA, narcolepsy, psychiatric sleep disorder, and periodic limb movement (PLM) disorder. Due to its length (176 items), the researchers concluded that the SDQ is not a practical instrument for use in screening the general population for common sleep disorders. The Pittsburgh Sleep Quality Index (PSQI) was developed to assist in measuring sleep quality, and, in turn, to alert physicians of the need to further evaluate their patients. This 24-item questionnaire is most appropriate for either measuring changes in sleep quality over time in patients, or in measuring differences in sleep quality between diseased groups, but does not aid in the diagnosis of a particular sleep disorder [10].

As such, no comprehensive screening questionnaire could be identified that was practical for use in primary care, and had undergone rigorous testing of its psychometric properties. Therefore, a screening tool for the most prevalent sleep disorders in adults — the Global Sleep Assessment Questionnaire (GSAQ) — was developed. The primary objectives of this study were to implement the GSAQ in primary care and sleep centers with the following aims: (1), to develop optimal criteria for identifying each target sleep disorder with the GSAQ; (2), to evaluate the psychometric properties (content validity, test–retest reliability, construct validity, sensitivity/specificity) of the GSAQ for each of these disorders; and (3), to evaluate the potential utility of the GSAQ as a screening tool for sleep disorders in primary care and sleep specialty centers.

2. Materials and methods

2.1. Overview

Eligible and interested patients who presented to either a sleep center or a primary care clinic with no previous diagnosis of a sleep disorder were enrolled in this multi-center study. All participants completed the GSAQ at least once and underwent a clinical evaluation by a board-certified sleep specialist.

The GSAQ is a self-administered screening tool which was designed for use by clinicians to aid in the diagnosis of sleep-related disorders. The questionnaire used in the current study consisted of several sections. The first section contained an 11-item GSAQ which used consistent four-item response options (e.g. 'never', 'sometimes', 'usually', 'always'), while the 19-item second section varied the response options and was more quantitative in nature (e.g. '0 times/week', 'more than 4 times/week'). All original items were developed specifically for this study to screen for and differentiate between the following seven disorders: insomnia disorders, insomnia associated with a mental disorder (IME), OSA syndrome, restless legs syndrome (RLS), PLM disorder, parasomnias, and shift work sleep disorder. Additionally, the study questionnaire contained two generic quality-of-life outcome measures: the SF-36, measuring the eight domains of physical functioning, social functioning, role limitations due to physical limitations, role limitations due to emotional functioning, mental health, vitality, pain, and general health perceptions [15]; and the Medical Outcomes Study (MOS) Sleep Problems Index I [16]. Lastly, the questionnaire included items assessing recent stressful events, presence of children under the age of 2 years in the home, employment status (i.e. day, night, or rotating shift) and demographic information: age, race, marital status, household income, and education level.

A clinical form was completed by the board-certified sleep specialist following each patient's clinical evaluation and/or diagnostic test (e.g. nocturnal polysomnography (PSG), or the multiple sleep latency test (MSLT)). This one-page form reported the presenting diagnosis, the final diagnosis (if made), and any methods used to confirm or rule out a diagnosis.

2.2. Procedures

Subjects in both the sleep centers and primary care centers were eligible for inclusion in the study if they were: at least 18 years of age, able to speak and read English, never previously diagnosed with a sleep disorder, willing to undergo a clinical evaluation conducted by a board-certified sleep specialist, and able to complete a self-administered questionnaire up to two times during the course of the study.

All sites received Institutional Review Board (IRB) approval for this study from a central IRB, and, when required, at the local level. Individuals presenting at participating sites were assessed for eligibility and provided with a description of the study. As necessary, sites placed advertisements in local newspapers, posted flyers in waiting rooms or clinics, and/or employed other appropriate recruitment strategies. All subjects were compensated monetarily for their time.

Five sleep centers agreed to recruit up to 45 eligible patients for the study through their clinic and/or community, of whom ten at each center were randomized to test–retest reliability testing using a random number list. The majority of sleep center participants were 'naïve' patients calling to schedule their first appointment at the sleep center (therefore, they had no prior diagnosis). The study imposed a maximum cap of ten subjects/diagnostic group per site in order to ensure as wide a diagnosis representation as possible. Once a site reached its enrollment cap for a presenting diagnosis, the investigator was instructed to cease enrolling patients presenting with this specific diagnosis.

Those participants who were *not* randomized to the test–retest group completed the questionnaire at their clinic visit. The questionnaire was self-administered by the patient in the waiting room prior to being evaluated by a sleep specialist. Participants who were randomized to the test–retest group self-administered the questionnaire in their homes, and then again 7–14 days later when they presented to the clinic.

Two primary care centers recruited up to 25 eligible patients for the study, all of whom completed the questionnaire twice for test–retest reliability testing. Patients recruited at the primary care center may or may not have had sleep complaints. After completing the questionnaire in the primary care clinic, the patient was referred to the affiliated sleep center to undergo a clinical evaluation by a board-certified sleep specialist. Each participant was also provided with a second copy ('retest') of the questionnaire to complete in his/her home 1–2 weeks later. The board-certified sleep specialist then evaluated each participant for potential sleep disorders. Following the clinical evaluation, the board-certified sleep specialist completed the clinical form (identical to that used in the sleep center portion of the study). All participating physicians in both the sleep centers and the primary care clinics were instructed to evaluate patients according to their usual standard of care.

2.3. Statistical methods

Scales for a total of ten categories were created: one for each of the seven specific sleep disorders, one for multiple disorders, one for 'other', and one for no sleep disorder (a confirmed diagnosis ruling out sleep disorder). Patients with more than one confirmed diagnosis were included in any analysis that included an applicable diagnosis.

Item responses were converted to a common 0–100 scale, with a higher score indicating greater likelihood for presence of the disorder. Domains of the GSAQ were then defined using a logistic regression procedure for each of the identified diagnoses. This procedure had four steps: univariate screen, multivariable modeling, construction of a domain score, and validation. The objective of the univariate screen was to identify which questionnaire items were associated with each type of confirmed diagnosis without adjustment for other questionnaire items. These items were then assessed together in a forced-fit multivariable logistic regression model predicting the target diagnosis. The objective of this analysis was to discard only those items with very low coefficients relative to other terms in the model. For the remaining terms, a second objective was to identify negative and positive predictors of the diagnosis based on the signs of coefficients. A third objective was to determine the relative weight of remaining terms by inspecting the size of coefficients of estimation associated with each term.

Each item remaining after the multivariable analysis was defined as a component of a domain score for a particular diagnosis. A candidate domain score was then developed based on the mean of the components, weighted for the direction of association (positive or negative) and the relative size of association. No missing values were allowed in the computation of these domain scores. As a sensitivity analysis, logistic regression was used to test whether the addition of individual component items (to the model containing only the candidate domain score) significantly improved the model. All final domains were also examined for face validity by inspecting the items used to develop the final domain score.

For each diagnosis, final domain scores were then entered in a logistic regression model predicting the likelihood of a specific sleep disorder. The area under the curve (AUC), a measure of overall accuracy, was calculated. A receiver operating characteristic (ROC) curve, plotting sensitivity against the false positive rate, was generated for each diagnosis.

Two cutpoints were selected for demonstration: the average domain score for patients known to have the disorder being examined, and the second was the average domain score for patients known not to have the disorder. Once optimal cutpoints for domain scores were characterized in an ROC analysis, other terms, representing population characteristics, were added to the logistic model.

Psychometric properties of the GSAQ items were evaluated. Content validity was established during the developmental process by soliciting input from board-certified experts in the field of sleep, and by comparing with other available sleep screeners. Additionally, simple frequencies for each item were computed to confirm that items were understood and that a range of responses was used.

Test–retest reliability was measured using the intra-class correlation coefficient (ICC), and was assessed by comparing the correlations between two administrations of the GSAQ for a subset of the study sample. A correlation of 0.70 or greater is considered acceptable [17].

Construct validity, using convergent and divergent validity, was evaluated by examining the Pearson correlation coefficients between the GSAQ scales and the SF-36 and MOS scales.

Known groups were established as disorder cohorts, and were based on confirmed clinical diagnosis. Validation was based primarily on predictive value, and the presence of a gold standard (the final diagnosis) eliminated the need for traditional known groups validity.

In order to test discriminant validity, a sensitivity–specificity analysis was conducted. Due to the inverse relationship between sensitivity and specificity, paired cutpoints were examined: one which maximizes sensitivity (while requiring a specificity of only 50%), and another which maximizes specificity (while requiring a sensitivity of only 50%).

3. Results

A total of 212 patients with a confirmed sleep diagnosis completed the study: 168 sleep center patients representing five sleep clinics and 44 primary care patients from two

Table 1
Demographic characteristics, overall and by type of center[a]

Characteristic	Overall ($N = 212$)	Sleep center ($N = 168$)	Primary care center ($N = 44$)	P value[b]
Gender				0.006
Male	102 (48.1)	89 (53.0)	13 (29.6)	
Female	110 (51.9)	79 (47.0)	31 (70.5)	
Ethnicity				0.14
White	160 (75.5)	123 (73.2)	37 (84.1)	
Non-white	52 (24.5)	45 (26.8)	7 (15.9)	
Education				0.001
Some college	158 (74.5)	134 (79.8)	24 (54.6)	
No college	54 (25.5)	34 (20.2)	20 (45.5)	
Income[c]				0.61
≤$49,999	93 (51.1)	75 (52.1)	18 (47.4)	
≥$50,000	89 (48.9)	69 (47.9)	20 (52.6)	
Child less than 2 years old in home[c]				0.30
Yes	16 (7.7)	11 (6.7)	5 (11.4)	
No	192 (92.3)	153 (93.3)	39 (88.6)	
Employment status[d]				
Full-time	111 (52.4)	90 (53.6)	21 (47.7)	0.49
Day shift	86 (40.6)	70 (41.7)	16 (36.4)	0.52
Night shift	26 (12.3)	21 (12.5)	5 (11.4)	0.84
Rotating shift	13 (6.1)	9 (5.4)	4 (9.1)	0.36
Homemaker	25 (11.8)	20 (11.9)	5 (11.4)	0.92
Retired	31 (14.6)	19 (11.3)	12 (27.3)	0.008
Unemployed	15 (7.1)	12 (7.1)	3 (6.8)	0.94
Major stressful event				0.18
Environment	1 (0.5)	1 (0.6)	0 (0.0)	
Family	14 (6.6)	8 (4.6)	6 (13.6)	
Health	6 (2.8)	5 (2.8)	1 (2.3)	
Personal	5 (2.4)	5 (2.8)	0 (0.0)	
Work	7 (3.3)	7 (4.2)	0 (0.0)	
None	179 (84.4)	142 (84.5)	37 (84.1)	
Number of caffeinated beverages/day				0.003[b]
0/day	49 (23.1)	43 (25.6)	6 (13.6)	
1–3/day	126 (59.4)	99 (58.9)	27 (61.4)	
4–7/day	28 (13.2)	23 (13.7)	5 (11.4)	
8–10/day	9 (4.2)	3 (1.8)	6 (13.6)	
More than 10/day	0 (0.0)	0 (0.0)	0 (0.0)	
Number of alcoholic beverages/day				0.47[b]
0/day	120 (56.6)	97 (57.7)	23 (52.3)	
1–3/day	55 (25.9)	45 (26.7)	10 (22.7)	
4–7/day	18 (8.5)	14 (8.3)	4 (9.1)	
8–10/day	10 (4.7)	6 (3.6)	4 (9.1)	
More than 10/day	9 (4.2)	6 (3.6)	3 (6.8)	
Age	45.0 ± 14.12 (18–87)[e]	44.1 ± 13.4 (18–76)[e]	48.5 ± 16.2 (18–87)[e]	0.065[f]
Weight	183.2 ± 58.06	187.6 ± 61.9	164.3 ± 33.6	0.018[f]
Body mass index (kg/m^2)	28.3 ± 8.34	28.9 ± 8.9	26.1 ± 4.9	0.053[f]

[a] Data represent numbers followed by percentage values in parentheses or means ± SD.
[b] From Chi-square test; P values represent the difference between sleep and primary care centers.
[c] Missing data accounts for numbers not adding to a total of $N = 212$.
[d] Patients could report more than one category.
[e] Figures in parentheses represent the range.
[f] From t-test.

primary care clinics. Table 1 displays the demographic characteristics of the overall study group, as well as separately reporting those of patients from the sleep and primary care centers. Overall, the majority of respondents were female and Caucasian, had at least some college education, reported an income of $49,999 or less, did not have a child less than 2 years old living in their home, and were employed full-time. The mean age of the study group was 45.0 years. The demographics for participants in sleep and primary care centers were comparable in most respects, however, the sleep center group had a higher percentage of male respondents ($P = 0.006$) and more people with a college education ($P = 0.001$), while more individuals recruited from the primary care center were retired ($P = 0.008$). As patients

were recruited through both clinical referrals and community advertisements, this study population is considered to be a mixed clinical/community sample.

3.1. Clinical characteristics

All sleep specialists evaluated the patients according to their usual standard of care. This resulted in between-site differences (e.g. in the number of patients referred for sleep laboratory testing) owing to the clinical judgement and usual practices of physicians at each site.

Table 2 contains the overall clinical characteristics for all patients, and from the sleep and primary care centers individually. While many patients presented with only one diagnosed sleep disorder (65.6%), a large percentage (28.3%) of patients presented with and were found to have multiple diagnoses. A commonly-encountered combination of diagnoses was RLS and PLM disorder. Of the patients who received multiple diagnoses, 52 patients presented with two, while 28 presented with symptoms of more than two. After confirmatory clinical evaluation, these numbers were reduced to 29 and 31, respectively. The most common test used to confirm a diagnosis was a PSG, followed by the MSLT. The diagnoses of OSA and PLMs were further confirmed by PSG in 62 and 43% of the patients, respectively. MSLTs were conducted in 12% of patients with sleep apnea and 11% of patients with parasomnias.

While 30% of the primary care cohort presented with no sleep disorder, over 40% of the primary care patients both presented with and were confirmed to have insomnia, followed by almost 16% with OSA. Approximately 11% were diagnosed with shift work disorder, and 18% had an 'other' sleep disorder. No confirmatory lab testing was performed on any of the primary care patients; rather, diagnoses were confirmed by clinical evaluation (see Table 2).

3.2. Reliability

Ninety-one participants (43%) completed a second 'retest' questionnaire. On average, the interval between test and retest was 12.6 days (12.3 days for participants at the sleep center, 12.9 days for participants in primary care). Acceptable test–retest reliability (indicated by ICC values) was found for primary insomnia (0.86), insomnia–mental (0.72), OSA (0.88), RLS (0.77), PLM (0.80), shift work (0.92), no-disorder (0.77), and multiple sleep disorder (0.75). Some of the less common diagnoses fell below the generally acceptable level, presumably due to an insufficient amount of variability in the responses [parasomnia (0.51), and other diagnoses (0.65)]. Further investigation reviewing these lower ICCs demonstrated that the ICC for parasomnia at the sleep centers was 0.63, while it was 0.34 for the primary care centers.

3.3. Validity

In general, higher correlations were found between diagnoses of similar types (convergent validity), and lower correlations were found between diagnoses that were more dissimilar (divergent validity; data not shown). For example, the scale assessing *PLM* was more highly correlated with *RLS* than with the scale indicating *primary insomnia* ($r = 0.57$ vs. 0.06). Certain scales also had strong negative correlations. For example, the *primary insomnia* scale had a correlation of $r = -0.68$ ($P = 0.0001$) with the *OSA* scale, and the *insomnia–mental* scale had a correlation of $r = -0.53$ ($P = 0.0001$) with the *no-disorder* scale. This implies that insomnia was not strongly associated with OSA, or that patients were not likely to be diagnosed with both insomnia and OSA. Additionally, patients with IME were not at all likely to respond similarly to those who have no sleep disorder.

Psychometric properties of the 11-item GSAQ were compared with the additional items included for validity testing purposes. The GSAQ was comparable with the other series of items in terms of reliability, validity, sensitivity, and specificity for determining the presence of a sleep disorder diagnosis. There were only slight decreases in the

Table 2
Clinical characteristics[a]

Diagnosis[b]	Overall confirmed ($N = 212$)	Confirmed diagnoses from sleep centers ($N = 168$)	Confirmed diagnoses from primary care centers ($N = 44$)
Primary insomnia (P)	77 (36.0)	59 (34.7)	18 (40.9)
Insomnia–mental (IME)	29 (13.6)	28 (16.5)	1 (2.3)
OSA	66 (30.8)	59 (34.7)	7 (15.9)
RLS	29 (13.6)	25 (14.7)	4 (9.1)
PLM	14 (6.5)	14 (8.2)	0 (0.0)
Parasomnias (P)	9 (4.2)	9 (5.3)	0 (0.0)
Shift work (SW)	26 (12.1)	21 (12.4)	5 (11.4)
Other sleep diagnoses	59 (27.6)	51 (30.0)	8 (18.4)
No sleep disorder	13 (6.1)	0 (0.0)	13 (29.5)
Multiple diagnoses			
2	29 (13.7)	21 (12.5)	8 (18.2)
>2	31 (14.6)	29 (17.3)	2 (4.5)

[a] Data represent numbers (n) followed by percentage values in parentheses.
[b] More than one diagnosis could have been recorded per patient.

sensitivity and specificity for some disorders (*insomnia*, *insomnia–mental*, *OSA*, *shift work*), following removal of a total of 19 items from the GSAQ. For instance, the sensitivity of the *OSA* scale decreased from 95% in the longer series of questions to 93% in the GSAQ.

Pearson correlations were calculated with the SF-36 (scale scores as well as mental and physical component scores). A higher score on the SF-36 indicates better functioning. Since a higher score on a GSAQ scale was designed to indicate presence of the disorder, those disorders that are more psychiatrically-based had a stronger negative correlation with both the mental health scale and the mental component score of the SF-36. For example, there was an observed correlation between the mental health scale and the *insomnia–mental* scale of 0.48 ($P = 0.0001$). As additional confirmation of the validity of the GSAQ, the *no-disorder* scale had a positive correlation with all the SF-36 scales, indicating that these patients self-reported better functioning and better overall sleep health.

Finally, the six-item MOS Sleep Problems Index I was also used to confirm the construct validity of the screening instrument. Items in the MOS scale assess trouble falling asleep, difficulties with sleep maintenance, respiratory problems, sleep adequacy, and daytime sleepiness. A higher score indicates a greater presence of the attribute (i.e. more difficulty falling asleep). There were moderate positive correlations between the MOS Sleep Problems Index I and the GSAQ with the *OSA* ($r = 0.37$; $P = 0.0001$), *PLM* ($r = 0.33$; $P = 0.0001$), as well as the *insomnia–mental* ($r = 0.49$; $P = 0.0001$) scales. The validity of the questionnaire is also upheld through the finding that those patients with no confirmed sleep disorder had a strong negative correlation with the MOS sleep scale of -0.68 ($P = 0.0001$).

3.4. Known groups validity

Table 3 displays scale scores for all patients, and for patients with each specific diagnosis. Patients confirmed with a given sleep disorder are by definition clinically different from all other patients participating in the study, and it is, therefore, possible to perform known-groups analyses segregated by each disorder cohort. Those patients clinically diagnosed with a disorder scored highest on the specific GSAQ scale for that disorder. For instance, the overall score for the primary insomnia scale was 59.8 and those patients diagnosed with *primary insomnia* and *insomnia–mental* scored well above the overall average (67.6 and 66.5, respectively). Disorders with similar clinical characteristics also demonstrated similar trends in scores: for example, the overall score on the RLS scale was 47.9, and patients diagnosed with *RLS* (65.1), *PLM* (63.2), and *parasomnias* (59.5) all scored above the overall score (see Table 3).

Fig. 1 shows the mental and physical component scores of the SF-36 by each disorder cohort for both primary and sleep centers. The *insomnia–mental* group reported the lowest mental component scores. The highest values for both mental and physical component scores were found in the *no-disorder* group, supporting the validity of the GSAQ. The *parasomnia* group reported high physical component scores, which is to be expected as patients with parasomnia do not typically report problems in their daily physical activities.

3.5. Discriminant validity

Table 4 displays the results of the sensitivity and specificity analyses. Cutpoint 1 was selected because it represents the mean score for individuals without the specific diagnosis in question, while cutpoint 2 represents the mean score for those individuals with that specific diagnosis. Therefore, by definition, the specificity for cutpoint 1 will be close to 50%, and the sensitivity for cutpoint 2 will approximate 50%.

The AUC value can be considered as the single best summary of sensitivity and specificity with 100% representing the ideal value. Lower AUCs were found for primary insomnia (72%) compared with some of the less prevalent disorders (e.g. parasomnia = 95%).

Data were also analyzed by adjusting for verification bias, sex, type of center (primary vs. sleep), and age. While verification bias and gender did not have a significant effect on the sensitivity and specificity findings, adjustments for age had an impact on the sensitivity for primary insomnia, reducing it from 79 to 57%.

Table 3
Sleep scale scores for patients, sleep and primary care centers combined[a,b]

Scale	Overall ($N = 240$)	Primary insomnia ($N = 76$)	Insomnia–mental ($N = 28$)	OSA ($N = 65$)	RLS ($N = 29$)	PLM ($N = 14$)	Parasomnias ($N = 9$)	Shift work ($N = 26$)
Primary insomnia (I)	59.8 ± 14.1	67.6 ± 10.9	66.5 ± 9.9	49.5 ± 13.3	63.9 ± 11.0	64.1 ± 13.0	69.8 ± 6.3	64.3 ± 9.2
Insomnia–mental (IME)	42.1 ± 14.3	46.4 ± 13.2	56.1 ± 14.4	41.7 ± 17.3	47.6 ± 13.2	45.5 ± 10.5	46.3 ± 8.6	46.8 ± 12.2
OSA	22.6 ± 19.0	16.0 ± 15.0	21.5 ± 18.3	40.7 ± 17.6	17.4 ± 15.1	21.7 ± 15.4	5.6 ± 3.2	22.0 ± 14.1
RLS	47.9 ± 15.3	49.4 ± 12.3	51.0 ± 13.1	41.3 ± 16.5	65.1 ± 14.3	63.2 ± 21.2	59.5 ± 8.9	48.4 ± 12.3
PLM	20.9 ± 19.3	18.9 ± 17.5	24.6 ± 18.2	24.2 ± 18.3	37.3 ± 20.8	46.8 ± 19.6	38.9 ± 14.6	22.2 ± 15.7
Parasomnias (P)	45.5 ± 10.9	45.2 ± 10.2	48.0 ± 11.0	44.1 ± 10.9	46.5 ± 10.0	48.0 ± 12.6	64.6 ± 7.2	39.4 ± 10.3
Shift work (SW)	14.7 ± 27.8	17.7 ± 32.0	19.5 ± 34.7	15.1 ± 27.4	13.1 ± 27.4	11.0 ± 22.0	11.9 ± 28.2	63.1 ± 40.3
None	60.7 ± 19.6	55.6 ± 17.5	45.4 ± 16.6	59.8 ± 20.0	60.5 ± 17.5	48.4 ± 16.7	51.9 ± 19.3	55.1 ± 17.1

[a] Means ± SD.
[b] Total number of patients with diagnosis; evaluable sample sizes varied with scale. All scale results combine both sleep and primary care centers.

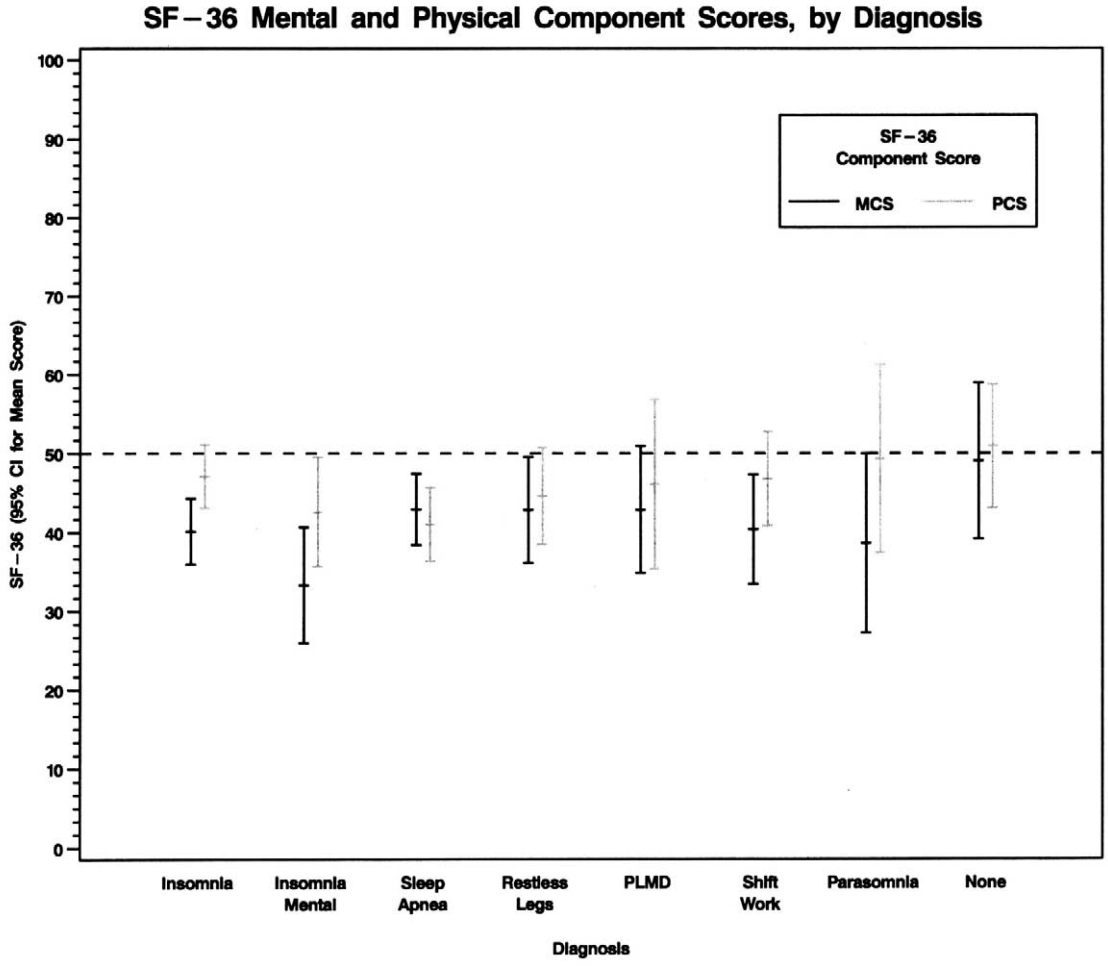

Fig. 1. SF-36 component score: (black line), MCS; (shaded line), PCS.

3.6. Comparison of results with other findings

Normative data for the SF-36 are available for the general US population [16], which is valuable in comparing our study population with a sample of 'healthy' individuals (i.e. those without a sleep disorder who are not presenting to a primary care clinic). The normative cohort scored slightly higher (representing better functioning) on all domains of the SF-36 when compared with both sleep center and primary care patients. Large differences between sleep center patients and normative data were found on nearly all scales, confirming a much lower self-reported level of functioning for sleep center patients. Significant differences were seen in the domains of role functioning (53.3 ± 41.6 vs. 80.96 ± 34.0), general health (58.6 ± 24.3 vs. 71.95 ± 20.34), energy and vitality (36.9 ± 22.3 vs. 60.86 ± 20.96) and social functioning (66.4 ± 29.0 vs. 83.26 ± 22.69). Additionally, Zammit et al. [18] found that, when compared with a control group, patients with insomnia had significantly ($P < 0.0001$) lower mean scale scores on all subscales of the SF-36, as well as on the MOS cognitive scale and both the Zung Anxiety and Depression scales [18]. Another recent study found that patients with OSA syndrome scored consistently lower than controls on the majority of SF-36 subscales [19]. These findings suggest that individuals with a sleep disorder report worse functioning and well-being, and that those from a primary care setting report comparable functioning and well-being as compared with published normative data.

4. Discussion

Optimally, the GSAQ could be used to screen for sleep disorders in a general primary care population. In addition, it may be useful to administer the GSAQ on a regular basis to ensure that no sleep disorder has developed. The specific scoring guide for the GSAQ evaluates each diagnosis separately, and sums item responses to designate whether a patient merits additional investigation into sleep disorders.

The clinical utility of a sleep screening tool depends on predictive values likely to be experienced under routine practice conditions. For a given sensitivity or specificity, the positive and negative predictive values associated with

Table 4
Sensitivity and specificity results, Form A GSAQ

Scale[a]	*N*	Frequency (%)	AUC (%)	Sensitivity (%)	Specificity (%)	False positives	False negatives
Primary insomnia (I)	177	34	72				
Cutoff 1: 56				79	57	50	13
Cutoff 2: 67				51	77	27	30
Insomnia–mental (IME)	179	13	78				
Cutoff 1: 41				83	51	77	4
Cutoff 2: 56				48	85	23	12
OSA	176	32	88				
Cutoff 1: 14				93	58	50	4
Cutoff 2: 41				44	95	6	32
RLS	194	14	84				
Cutoff 1: 46				96	50	83	1
Cutoff 2: 65				48	90	17	14
PLM	205	7	84				
Cutoff 1: 20				93	52	91	1
Cutoff 2: 47				43	93	14	8
Parasomnias (P)	208	4	95				
Cutoff 1: 45				100	49	101	0
Cutoff 2: 65				33	98	3	6
Shift work (SW)	213	12	88				
Cutoff 1: 8				77	88[b]	23	6
Cutoff 2: 63				65	95	9	9
None	210	6	84				
Cutoff 1: 58				92	56	87	1
Cutoff 2: 82				69	91	18	4

[a] Cutoff 1 is the score average for the group without diagnosis; cutoff 2 is the score average for the group with diagnosis.
[b] Forty-six percent of patients without diagnosis had a score 1 point below the mean.

a particular cutpoint are influenced strongly by the prevalence of the disorder in the screening population, as well as by differences in patient characteristics. This means a test with a given sensitivity or specificity can provide different predictive values in different settings. As an example, when the prevalence of a target disorder is relatively high, such as that which might occur in a referred population, positive predictive values are likely to be higher than if the same sensitivity were applied in a low prevalence setting. Conversely, positive predictive values are more highly influenced by test specificity when the prevalence of target disorder is expected to be relatively low, such as that which might occur in a naïve population.

Through the course of analyses, several generalizations were made about the course of treatment for certain disorders. For example, a patient with insomnia is more likely to be treated without referral or further testing by a primary care physician. Therefore, it is necessary to have a high enough specificity which minimizes the number of false positives (i.e. patients being diagnosed with insomnia who do not actually have the disorder). For example, a false positive GSAQ result, indicating insomnia, would not typically result in costly interventions (and, for example, a lower AUC may be acceptable). That, however, might not be the case for OSA, where in most instances, a PSG or another costly laboratory test may be the next step in confirming a diagnosis. Therefore, we need to ensure that the specificity is high. While the GSAQ is not intended in any way to supplant the clinical expertise of a physician, but rather to indicate areas which merit further investigation, it is appropriate to set cutpoints at a level which would reduce the likelihood of false positives while still indicating those patients who may be 'borderline' cases of the disorder. Conversely, for certain sleep disorders which would result in a referral to a sleep specialist (PLM or RLS), it is important that the instrument be able to indicate the possible presence of the disorder. In the instances when follow-up or corroborative testing is expensive or harmful, specificity may be considered more important than sensitivity. The more modest the immediate intervention is, or if the objective is to screen a high-risk population, the more useful it is to maximize sensitivity over specificity.

Due to between-site differences in the prevalence of confirmatory testing and patient demographics, data were also analyzed for verification bias and potential confounding effects or interactions with sex, type of center, and age. The association between the scales and the diagnoses as measured by an odds ratio from a logistic regression model remained relatively unchanged when these variables were included in the model as potential covariates. Additionally, the sensitivity and specificity did not differ substantially when they were recomputed stratified by subgroup. For these analyses, only age had a significant effect on the sensitivity and specificity findings for insomnia. Therefore, although the sample sizes were small, it is possible that the screening tool will be more effectively utilized with younger

patients, that is, those under the age of 65. Since the majority of sleep disorders are typically present prior to the age of 65 years, we do not feel that this limits the power of our tool.

Some of the less common diagnoses fell below the generally acceptable level of 0.70 for reliability, presumably due to an insufficient amount of variability in the responses. As seen in the results, the lower ICC values at primary care sites confirm that the lack of variability with a concentration at the low end of the scales is largely responsible for the low ICCs for the uncommon diagnoses. Ideally, the ICC would also be computed for only the subset of patients with the confirmed diagnosis, but too few patients with these diagnoses were available in the test–retest cohort. For example, ICC values utilizing *only* those patients who had been confirmed with a diagnosis of parasomnia were never computed because the sample size was very low.

The positive predictive value is impacted significantly by the prevalence of a disorder (those more common in the population are more likely to be positively predicted). Therefore, we believe that the positive predictive value for the general US population is less useful than the separate predictive values for patients presenting at sleep centers and patients presenting in primary care settings. Testing of the correlations between scales was performed, however, as each scale is a sleep disorder diagnosis and the instrument is designed to discriminate between them, we would not expect any of the scales to be predictive or associated with other disorders.

Several study limitations should be considered when interpreting our findings. Due to the large number of sleep disorders of interest in this current study, several of the diagnostic groups had small number of confirmed diagnoses (i.e. parasomnias, PLM). Therefore, for those groups, we can only make observations about what may be appropriate sensitivity and specificity cutpoints. Further work on evaluating the positive predictive value of the GSAQ should be carried out. Additionally, we did not conduct traditional known-groups analyses in order to assess validity. Finally, the study design did not mandate diagnostic testing for confirmation of each diagnoses. While this is more closely aligned with the actual implementation setting (i.e. 'usual care'), it also does not provide a gold standard for obtaining and confirming a given diagnosis. In general, primary care patients reported a higher than expected prevalence of sleep disorders. While a portion of the study utilized patients from a primary care setting, the GSAQ needs to be implemented in a primary care setting in order to further determine its utility. It would be invaluable to the structure of the GSAQ to obtain feedback from primary care physicians who might be implementing it in their practice. Nonetheless, we believe that the GSAQ is a practical and simple tool to help screen for sleep disorders. In contrast to other sleep questionnaires, the GSAQ is a short questionnaire that addresses a number of common sleep disorders, rather than just focusing on one specific disorder. The GSAQ may be used with other questionnaires that measure sleep quality or symptoms to further understand a patient's sleep disorder.

We recommend the use of the 11-item GSAQ as a screening tool. Further discussions and feedback from primary care physicians would aid in the refinement of the screener and practical institution of the survey.

Acknowledgements

The authors would like to thank Elisabeth Warren, whose help in all aspects of the study has been invaluable, and Dave Miller and Michelle Pritchard for their statistical expertise. Recruitment would not have been possible without the contributions of Syed Akbarullah M.D., Laurie DeSimone, Minisha Kochar, Gail Koshorek, John Pinto M.D., Kathleen Rice Ph.D., Sharon Schutte M.D., Candida Sherrill, and Darlene Steljes at our study sites.

Appendix A. Sample questionnaire items

During the past 4 weeks, how often did you have difficulty falling asleep, staying asleep, or feeling poorly rested in the morning?

1. Never
2. Sometimes
3. Usually
4. Always

During the past 4 weeks, how often did you hold your breath, have breathing pauses, or stop breathing in your sleep?

1. Never
2. Sometimes
3. Usually
4. Always

During the past 4 weeks, how often did you have repeated rhythmic leg jerks or leg twitches during your sleep?

1. Never
2. Sometimes
3. Usually
4. Always

A full copy of this questionnaire can be obtained through the senior author.

References

[1] Wake up America: a national sleep alert, Report of the National Commission on Sleep Disorders Research 1993.
[2] US Bureau of the Census, 1993.
[3] Young T, Palta M, Dempsey J, Skatrud J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993;328:1230–1235.
[4] Stepanski E, Zorick F, Roehrs T, Roth T. Effects of sleep deprivation on daytime sleepiness in primary insomnia. Sleep 2000;23(2):215–219.
[5] Kuppermann A, Lubeck DP, Mazonson PD, et al. Sleep problems and

their correlates in a working population. J Gen Intern Med 1995;10:25–35.

[6] Hatoum HT, Kong SX, Kania CM, et al. Insomnia, health-related quality of life and health care resource consumption: a study of managed care organization enrollees. Pharmacoeconomics 1998;14(6):629–637.

[7] Kryger MH, Roos L, Delaive K, Walld R, et al. Utilization of health care services in patients with severe obstructive sleep apnea. Sleep 1996;19(Suppl 9):S111–S116.

[8] Netzer NC, Stoohs RA, Netzer CM, et al. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131(7):485–536.

[9] Douglass AB, Bornstein R, Nino-Murcia G, et al. The Sleep Disorders Questionnaire. I. Creation and multivariate structure of SDQ. Sleep 1994;17(2):160–167.

[10] Buysse DJ, Reynolds CF, Monk TH, Berman SR, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193–213.

[11] Baumel MJ, Maislin G, Pack AI. Population and occupational screening for obstructive sleep apnea: are we there yet? Am J Respir Crit Care Med 1997;155:9–14.

[12] Hallett M, Burden S, Stewart D, et al. Sleep apnea in end-stage renal disease patients on hemodialysis and continuous ambulatory peritoneal dialysis. ASAIO J 1995;41:M435–M441.

[13] Pradhan PS, Gliklich RE, Winkelman J. Screening for obstructive sleep apnea in patients presenting for snoring surgery. Laryngoscope 1996;106:1393–1397.

[14] Le Bon O, Verbanck P, Hoffmann G, et al. Sleep in detoxified alcoholics: impairment of most standard sleep parameters and increased risk for sleep apnea, but not for myoclonias — a controlled study. J Stud Alcohol 1997;58:30–36.

[15] Ware JE, Snow KK, Kosinski M, et al. SF-36 health survey manual and interpretation guide. Boston, MA: New England Medical Center, 1993.

[16] Stewart AL, Sherbourne CD, Hays RD, et al. Summary and discussion of MOS measures. In: Stewart AL, Ware Jr JE, et al., editors. Measuring functioning and well-being: the medical outcomes study approach, Durham, MC: Duke University Press, 1992. pp. 345–371.

[17] Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of review scientific criteria. Clin Ther 1996;18:979–992.

[18] Zammit GK, Weiner J, Damato N, Sillup GP, McMillan CA. Quality of life in people with insomnia. Sleep 1999;22(Suppl 2):S379–S385.

[19] Gliklich RE, Taghizadeh F, Windelman JW. Health status in patients with disturbed sleep and obstructive sleep apnea. Otolaryngol Head Neck Surg 2000;122(4):542–546.