

Evaluating Sleepiness-Related Daytime Function by Querying Wakefulness Inability and Fatigue: Sleepiness-Wakefulness Inability and Fatigue Test (SWIFT)

R. Bart Sangal, M.D., F.A.A.S.M.

Sleep Disorders Institute, Sterling Heights, MI; Oakland University William Beaumont School of Medicine, Rochester, MI

Study Objectives: Routine assessment of daytime function in Sleep Medicine has focused on “tendency to fall asleep” in soporific circumstances, to the exclusion of “wakefulness inability” or inability to maintain wakefulness, and fatigue/tiredness/lack of energy. The objective was to establish reliability and discriminant validity of a test for wakefulness inability and fatigue, and to test its superiority against the criterion standard for evaluation of sleepiness—the Epworth Sleepiness Scale (ESS).

Methods: A 12-item self-administered instrument, the Sleepiness-Wakefulness Inability and Fatigue Test (SWIFT), was developed and administered, with ESS, to 256 adults \geq 18 years of age (44 retook the tests a month later); consecutive patients with symptoms of sleep disorders including 286 with obstructive sleep apnea ([OSA], apnea-hypopnea index \geq 5/h sleep on polysomnography [PSG]), 49 evaluated with PSG and multiple sleep latency test for narcolepsy and 137 OSA patients treated with continuous positive airway pressure (CPAP).

Results: SWIFT had internal consistency 0.87 and retest intraclass coefficient 0.82. Factor analysis revealed 2 factors—general wakefulness inability and fatigue (GWIF) and driving wakefulness inability and fatigue (DWIF). Normal subjects differed from patients in ESS, SWIFT, GWIF, and DWIF. SWIFT and GWIF (but not DWIF) had higher area under ROC curve, Youden’s index, and better positive and negative likelihood ratios than ESS. ESS, SWIFT, GWIF, and DWIF improved with CPAP. Improvements in SWIFT, GWIF, and DWIF (but not ESS) were significantly correlated with CPAP compliance.

Conclusions: SWIFT is reliable and valid. SWIFT and its factor GWIF have a discriminant ability superior to that of the ESS.

Keywords: Sleepiness, wakefulness inability, fatigue, obstructive sleep apnea, Epworth Sleepiness Scale

Citation: Sangal RB. Evaluating sleepiness-related daytime function by querying wakefulness inability and fatigue: Sleepiness-Wakefulness Inability and Fatigue Test (SWIFT). *J Clin Sleep Med* 2012;8(6):701-711.

Sleepiness is a commonly reported public health problem. In a 2002 National Sleep Foundation (NSF) poll,¹ 7% reported sleepiness almost every day and another 9% a few days a week, for a total of 16%. Seventeen percent reported having dozed off while at the wheel of a vehicle, and 1% reported having an accident because they dozed off or *were too tired* (emphasis added). In other words, accidents were attributed to either dozing off or being tired. According to the National Highway Traffic Safety Administration (NHTSA),² “NHTSA data indicate in recent years there have been about 56,000 crashes annually in which driver drowsiness/fatigue (emphasis added) was cited by police. Annual averages of roughly 40,000 nonfatal injuries and 1,550 fatalities result from these crashes.” Sleep apnea patients report not just daytime sleepiness, but also being tired, fatigued, or having a lack of energy, and these complaints may be more frequent than sleepiness in sleep apnea.³ Sleepiness (and its adverse effect of auto accidents) may be multifactorial, with elements of inability maintaining wakefulness when necessary or desired (wakefulness inability), tendency to doze off in soporific circumstances, and fatigue/tiredness/lack of energy. However, the specialty of Sleep Medicine has tended to focus rather exclusively on routine evaluation of daytime function by measuring sleepiness defined as tendency to fall asleep in soporific circumstances (using the multiple sleep latency test [MSLT]⁴

BRIEF SUMMARY

Current Knowledge/Study Rationale: There is not a questionnaire instrument to measure wakefulness inability or difficulty staying awake in situations where staying awake is desirable, or one that simultaneously addresses symptoms related to pathological sleepiness and to fatigue/tiredness/lack of energy in sleep disorders patients. Conceivably, what could be better than being able to fall asleep when one wants to (low MSLT [Mean Sleep Latency Test] and even high ESS [Epworth Sleepiness Scale]), but be able to stay awake when one wants to and feel refreshed (not tired) during the day?

Study Impact: A reliable and valid self-rating instrument (Sleepiness-Wakefulness Inability and Fatigue Test or SWIFT) was created and shown to be superior to the criterion standard for sleepiness (ESS) with regard to specificity, sensitivity and discriminate ability. It should be added to the ESS in evaluating daytime consequences of sleep disorders.

and the Epworth Sleepiness Scale [ESS]^{5,6}), to the exclusion of wakefulness inability (difficulty maintaining wakefulness) and fatigue, in sleep disordered patients.

The MSLT defines sleepiness as the ability to fall asleep in a dark room when asked to do so. According to a review paper to establish Standards of Practice for the clinical use of the MSLT and MWT (Maintenance of Wakefulness Test), “the wide range in MSL makes it difficult to establish a specific threshold value for excessive sleepiness or to discriminate

patients with sleep disorders from non-patients.”⁷ Further, “the MSL change between pre- and post-treatment for an individual is probably meaningful, although comparison of these data with the normative values is not helpful.”⁷ This may be related to the use of a behavior (the ability to fall asleep quickly when lying down in a dark room) that may be a desirable and adaptive trait rather than abnormal state. The MWT^{8,9} sought to correct this by asking subjects to try and stay awake in a dimly lit room. However, staying awake sitting in a dimly lit room doing nothing is not particularly advantageous (as opposed to staying awake when sitting in a dimly lit car and driving). The MSLT and MWT measure different abilities,^{10,11} and treatment may improve “wakefulness inability” (the MWT) more than “sleep tendency” (MSLT).¹² Both the MSLT and MWT require a large investment in time and resources.

There are self-rating questionnaire instruments that aim to measure sleepiness, the most commonly used being the ESS. The ESS queries for tendency to fall asleep in a variety of circumstances, often soporific. Sanford et al.¹³ have reported the distribution of the ESS. In their sample of normal subjects, median ESS was 7-8, and 30.7% of normal subjects without insomnia reported an ESS score ≥ 10 , the widely used cutoff for abnormal excessive sleepiness. It has been shown that the ESS does not measure the same ability as the MWT.^{14,15} In patients who are severely sleepy on the MWT, the ESS was insensitive to the level of sleepiness as measured by the MWT. The ESS may¹⁶ or may not¹⁷ be correlated with the MSLT. A sample of 10,000 subjects with 71% response rate showed no correlation between the ESS and the adverse consequence of automobile accidents, although there was a correlation between dozing off while stopped in traffic (item 8 on ESS) and automobile accidents.¹⁸

Sleep disordered patients report fatigue, lack of energy, and tiredness in addition to sleepiness.³ Fatigue, tiredness and lack of energy are largely interchangeable terms, as suggested by Merriam-Webster Dictionary’s¹⁹ definitions of fatigue as “weariness or exhaustion from labor, exertion, or stress”, and tired as “drained of strength and energy: fatigued often to the point of exhaustion.” Although these symptoms may be separable from sleepiness/wakefulness inability (wakeful being defined by Merriam-Webster Dictionary as “not sleeping or able to sleep”), it is not clear that sleep disordered patients, the general public and the NSF,¹ or the police and the NHTSA,² can separate these symptoms clearly.

Conceivably, what could be better than being able to fall asleep when one wants to (low MSLT and even high ESS), but be able to stay awake when one wants to and feel refreshed (not fatigued) during the day? Is the ability to fall asleep easily a pathological problem or an adaptive ability?²⁰ This leads to the question of whether we should be querying instead the ability to stay awake when desired, along with fatigue.

There does not seem to be a questionnaire instrument to measure wakefulness inability or difficulty staying awake in situations where staying awake is desirable. Although fatigue inventories such as the 83-item Multidimensional Fatigue Inventory (MFI)²¹ and its 30-item short form (MFSI-sf)²² exist, there is no single short questionnaire that simultaneously addresses symptoms related to pathological sleepiness and to fatigue/tiredness/lack of energy in sleep disordered patients.

Thus, these other domains of sleep disorder complaints are not routinely queried or measured.

The hypothesis was that a self-rating instrument for assessing wakefulness inability and fatigue can be created that is reliable (with good internal consistency and test-retest reliability in normal subjects) and valid (with good ability to discriminate between normal individuals and sleep disordered patients, and to show improvement with treatment of sleep disorders such as obstructive sleep apnea [OSA]), and that such a test incorporating wakefulness inability and fatigue is superior to the criterion standard for sleepiness (ESS) with regard to specificity, sensitivity, and discriminant ability.

METHODS

A 12-item questionnaire (Sleepiness-Wakefulness Inability and Fatigue Test, or SWIFT) was developed. Subscale A has 6 questions related to difficulty staying awake/wakefulness inability in different situations that might affect performance or cause adverse consequences; subscale B has 6 questions related to fatigue, tiredness or lack of energy in different situations that might affect performance or cause adverse consequences, all answered on a 4-level (scored 0-3) Likert scale. Items were prepared by the author based on apparent face validity, with the inclusion of more than one item related to driving. The SWIFT is shown in **Table 1**.

Normal Subjects

After obtaining approval from the Wayne State University Human Investigations Committee, adult subjects (age ≥ 18 years) were recruited over a period of 10 weeks by means of a group e-mail to medical students at Wayne State University as well as by personal solicitation of subjects in public places such as malls and parks. After reading an informational sheet, they were asked to fill out questionnaires seeking their gender, age, educational level, occupation, race, height, weight, medical/psychiatric problems, medicines taken, sleep habits, and presence or absence of sleep symptoms including snoring, observed or perceived apneic episodes in sleep, insomnia, fatigue, and sleepiness. They were also asked to complete the SWIFT and the ESS. Subjects willing to be contacted again in a month to retake the questionnaire were asked for contact information, and were contacted after a month to again complete the questionnaire.

A total of 403 subjects filled out the questionnaire. Subjects with incomplete questionnaires (49) were excluded. In order to examine the normal range of sleepiness, wakefulness inability, and fatigue, it was decided to exclude subjects with issues known to affect sleepiness, wakefulness inability, and fatigue, such as CNS-active or psychotropic medicines (53), CNS disorders (3), untreated depression (19), and history of observed/perceived apneic episodes in sleep (23). This left 256 normal subjects (87 male, 169 female; age range 18-92 years; 190 White, 26 Black, 5 Hispanic, 11 Asian American, 6 South Asian American, 18 Other; 4 with less than high school education, 23 high school graduates, 52 with some college, 101 with college degrees, and 76 with graduate degrees; National Statistics socioeconomic classification included: 10 higher professional or managerial, 50 lower professional or managerial, 26 intermedi-

Table 1—Sleepiness-Wakefulness Inability and Fatigue Test (SWIFT)

Name: _____

Date: _____

If this is the first time you are filling this out for this study, answer keeping in mind the last one month. Mark symptoms present only if they have been present for at least one month. If you have filled out this questionnaire before for this study, answer based on the period of time since you last filled this out. Be sure to answer every question to the best of your ability. This questionnaire refers to your usual way of life in recent times. Even if you have not done some of these things recently try to work out how they would have affected you. Use the following scale to choose the most appropriate number for each situation:

A How much of a problem is it to stay awake during the day (or your usual wake period if you sleep during the day)?

	Not at all	Just a little	Pretty Much	Very Much
1 Struggling to stay awake during the day	0	1	2	3
2 Difficulty staying awake while driving	0	1	2	3
3 Difficulty staying awake stopped at a traffic signal	0	1	2	3
4 Difficulty staying awake at work or while doing tasks	0	1	2	3
5 Difficulty staying awake while reading or studying	0	1	2	3
6 Difficulty staying awake in social situations	0	1	2	3

B How much of a problem has fatigue, tiredness or lack of energy been for you?

	Not at all	Just a little	Pretty Much	Very Much
1 Feeling tired when at work or while doing tasks	0	1	2	3
2 Lack of energy during social situations	0	1	2	3
3 Struggling with fatigue during the day	0	1	2	3
4 Feeling tired while reading or studying	0	1	2	3
5 No energy to do tasks that do not absolutely have to be done	0	1	2	3
6 Difficulty driving because of fatigue	0	1	2	3

© R. Bart Sangal, M.D., 2011. Permission is given to copy for use in medical care of personal patients.

The total score is the sum of the scores of all 12 items (A1-6, B1-6), with a maximum possible score of 36.

ate occupations, 3 small employers and own account workers, 12 lower supervisory and technical, 4 routine occupations, 36 retired or unemployed, and 115 students. Forty-four of them retook the SWIFT and ESS a month later.

Determining Reliability

To determine internal consistency, Cronbach α was calculated for the SWIFT (and the ESS) using data from the normal subjects. To determine test-retest reliability, intraclass coefficients were calculated using the normal subjects with test and retest data. If these tests showed good reliability (Cronbach α and intraclass coefficient > 0.8), factor analysis with varimax rotation was performed using SWIFT data.

Additional Analyses in Normal Subjects

Correlation between ESS and SWIFT was calculated. Males and females were compared with regard to age, BMI, time in bed, SWIFT and ESS, as were subjects who completed the questionnaire again and those who did not. Correlations were calculated between SWIFT and ESS on the one hand, and age, time in bed and BMI on the other. If there was a significant correlation between SWIFT and age, normal subjects were divided into two age groups to determine if the correlation persisted. If it did not, factor analysis and correlations were performed again by age group.

Correction for Multiple Statistical Analyses

In order to correct for multiple statistical analyses, the false discovery rate method was applied to primary but not to condi-

tional analyses (analyses performed only as a result of another statistically significant analysis).²³ This method rank orders the p-values of the analyses. For k analyses, one p value of 0.05/k was accepted as significant, one p value of 0.05/(k-1) was accepted as significant, and so on.

Determining Validity

After establishing good reliability for the SWIFT, validity was determined for SWIFT and its factors using data from normal subjects and sleep disordered patients, and SWIFT and its factors were compared with ESS to establish superiority.

Sleep Disordered Patients

All new patients presenting with sleep disorder symptoms to an AASM accredited Sleep Disorders Center during a 15-month period were administered the SWIFT and ESS at the time of the initial evaluation. If appropriate, they underwent a polysomnography (PSG) to evaluate for OSA using American Academy of Sleep Medicine (AASM) scoring criteria,²⁴ or a PSG with MSLT to evaluate for Narcolepsy. New patients who had previously been evaluated/treated for OSA anywhere were excluded from analysis. All patients with significant OSA (apnea-hypopnea index (AHI) ≥ 15 /h sleep, or ≥ 5 /h sleep with comorbid sleepiness, hypertension, or cardiovascular disease, evaluated as new patients over the 15-month period, were offered standard treatment. Patients opting for continuous positive airway pressure (CPAP) were administered a PSG with CPAP titration, and were prescribed CPAP at the optimum determined pressure. At a follow-up office visit between 1 and 3 months after CPAP

prescription, they were again administered the SWIFT and ESS, and CPAP compliance data were downloaded if available.

Data were available for 286 adult subjects (age ≥ 18 years, 192 males, 94 females) who presented with sleep disorder symptoms and had documented OSA (AHI ≥ 5 /h sleep). After excluding subjects with AHI ≥ 5 /h sleep on the PSG preceding the MSLT (who were counted among the adult OSA subjects), and patients who were administered the MSLT on CPAP, data were available for 49 adult subjects (17 males, 32 females) who were administered PSG with MSLT for suspicion of narcolepsy (because of unexplained sleepiness with no clinical evidence of OSA, off CNS-active medicines, including psychotropic medicines, for five half-lives). These 49 subjects were independent of and not a subset of the 286 OSA patients. Repeat ESS and SWIFT and compliance data from follow-up visits after CPAP initiation were available for 137 adult OSA patients (98 males, 39 females).

Determining Discriminant Validity

To determine discriminant validity, SWIFT (and SWIFT factor) and ESS scores were compared between the normal subjects and the OSA patients, as well as between normal subjects and patients evaluated for suspicion of narcolepsy. To further determine discriminant validity, SWIFT (and SWIFT factor) and ESS scores were compared in OSA patients before and after CPAP treatment, and the number of patients with abnormal SWIFT (and SWIFT factor) and ESS scores before and after CPAP treatment were compared. Correlations were calculated between compliance and improvement in ESS, SWIFT, and SWIFT factors identified by factor analysis.

Statistics of Diagnostic Tests

A diagnostic test identifies 2 groups: those with the disorder and those without the disorder. The sensitivity (also called true-positive rate) of a test is the probability of a positive test in the disordered population, whereas the specificity is the probability of a negative test in a disorder-free population; and the value (1-specificity) is also called the false-positive rate. The positive predictive value is the probability of a subject with a positive test having the disorder. The negative predictive value is the probability that a subject with a negative test does not have the disorder. Sensitivity and specificity are not affected by prevalence of the disorder, whereas positive and negative predictive values are affected. This means sensitivity and specificity can be accurately calculated when using a normal sample and a sample of disordered subjects, but predictive values cannot (they require a population sample for accurate calculations). A test with higher sensitivity and specificity than another is the superior test. However, a test may have higher sensitivity but lower specificity than another, or vice versa. Therefore, comparing 2 tests requires combining specificity and sensitivity. The likelihood ratio of a positive test or positive likelihood ratio (ρ_+) is the ratio of the probability of a positive test in a disordered subject (true-positive rate) to the probability of a positive test in a normal subject (false-positive rate), and is calculated as [sensitivity/(1-specificity)]. The likelihood ratio of a negative test or negative likelihood ratio (ρ_-) is the ratio of the probability of a negative test in a disordered subject to the probability of a negative test in a normal subject, calculated as [(1-sensitivity)/specificity]. Both likelihood ratios may range from 0 to α .

A positive likelihood ratio < 1 indicates a useless test, as does a negative likelihood ratio > 1 . With a diagnostic test based on a continuously measured variable, a decision or cutoff threshold allows sensitivity and specificity to be combined into the Youden's index γ , which is the true positive rate minus the false positive rate, calculated as [sensitivity-(1-specificity)] (also written as [sensitivity + specificity -1]). A perfect test (with sensitivity and specificity of 1) results in a Youden's index of 1, whereas a useless test has a Youden's index of 0. When the cutoff threshold is increased, the proportions of both true positives (sensitivity) and false positives (1-specificity) will increase. The receiver operating characteristic (ROC) is a graph of sensitivity against (1-specificity). A perfect test has an area under the ROC (AUC) of 1, a useless test has an AUC of 0.5. Bewick et al. have written an excellent but concise and simple discussion of these tests.²⁵ Since neither specificity nor sensitivity are affected by prevalence of the disorder, therefore positive and negative likelihood ratios, Youden's index and AUC are also not affected by prevalence when they are applied to population-based samples.

Determining Test Superiority

To determine which test is superior in discriminating normal subjects from sleep disordered patients, the AUC for the 2 tests can be compared. Visually, if the ROC for one test is entirely within the ROC of another test, then the second test seems certainly superior. Confidence intervals can be obtained for the AUC and statistical comparisons performed between the AUC for 2 tests using various nonparametric and binormal methods.^{26,27} A nonparametric distribution and correlated ROCs were assumed for this report. Different methods for calculating confidence intervals of Youden's index and likelihood ratios also exist, and a general method based on constant χ^2 boundaries was used for this analysis.²⁸ However, since the AUC is not dependent on a cutoff threshold, and diagnostic decisions are based on cutoff thresholds, a test may have a smaller AUC yet be more suitable than another, and the AUC is not a suitable measure of diagnostic excellence. Youden's index is a more suitable measure of diagnostic superiority.²⁹ Although Youden's index is a good single summary measure of comparison between two tests, positive and negative likelihood ratios are an even better test of superiority.³⁰ If test A has positive likelihood ratio greater than that for test B, and negative likelihood ratio lesser than that for test B, then test A is superior overall to test B. AUC, Youden's index, and positive and negative likelihood ratios (along with confidence intervals) were calculated for the ESS, SWIFT and its factors, using data from the normal subjects and OSA patients, as well as data from the normal subjects and patients evaluated for suspicion of narcolepsy. The AUC for SWIFT and ESS were compared. If there was a significant difference in favor of SWIFT, then the AUC for its factors were also compared with the AUC for ESS.

To determine which test is superior in showing improvement with treatment, effect sizes may be used. However, it is more important clinically to have a cutoff score (such as mean + 1 SD), above which the test is considered high and below which it is considered normal, and to show superiority in conversion of patients from abnormal scores before treatment to normal scores after treatment, therefore χ^2 analyses were also performed.

Additional Analyses in Sleep Disordered Patients

Correlation coefficients were calculated between ESS and SWIFT and its factors on the one hand, and sleep efficiency, arousal index, periodic limb movement arousal index (PLMAI), and AHI and lowest oxygen saturation (for OSA patients), mean sleep latency (MSL) and sleep onset REM periods [SOREMPS] (for patients evaluated for narcolepsy), with corrections for false discovery rate for multiple tests.

RESULTS

Reliability

Cronbach α using data from the 256 normal subjects was 0.87 for SWIFT and 0.80 for ESS. Upon retest, the intraclass correlation coefficient for SWIFT was 0.82 ($p < 0.001$), and for ESS 0.91 ($p < 0.001$).

Factor Analysis

Factor analysis of SWIFT with varimax rotation revealed 2 factors: Factor 1 (36% of variance) included 9 items (A1, A4, A5, A6, B1, B2, B3, B4, B5), and was called general wakefulness inability and fatigue (GWIF) based on the generality of the items. Factor 2 (20% of variance) included 3 items (A2, A3, B6), and was called driving wakefulness inability and fatigue (DWIF) based on these items being related to driving. **Table 2** gives the factor loadings.

Additional Analysis of Normal Subjects

ESS was correlated with SWIFT ($r = 0.64$, $p < 0.001$). There was no difference between males and females in age, BMI, hours in bed, SWIFT, or ESS. Those who completed the questionnaires again had a lower BMI (24.1 vs. 26.6, equal variances not assumed, $t = 2.9$, $df = 68.6$, $p = 0.005$) than those who did not, but there were no other significant differences. After

the false discovery rate correction, there were significant negative correlations between age and SWIFT ($r = -0.25$, $p < 0.001$) as well as ESS ($r = -0.14$, $p = 0.024$). There were no other significant corrected correlations.

Upon dividing the subject group into young adults (ages 18-45, $n = 188$) and middle-aged to older adults (age > 45 , $n = 68$), ESS and SWIFT were no longer correlated with age in either group. **Table 3** gives the measures of central tendency and dispersion for age, hours in bed, BMI, SWIFT, ESS, and the GWIF and DWIF factors for the 188 young adults and 68 middle-aged to older adults; the 85th percentile generally corresponds very closely to mean + 1 SD, and 95th percentile to mean + 2 SD.

Table 2—Factor analysis matrix of SWIFT

	Factor 1	Factor 2
B4	0.79	0.25
B3	0.77	0.19
B1	0.74	0.02
B2	0.73	0.09
A5	0.73	0.21
B5	0.72	0.04
A1	0.69	0.22
A4	0.56	0.10
A6	0.46	0.32
A2	0.04	0.90
A3	0.15	0.84
B6	0.26	0.80

Rotated component matrix for normal subjects by age group: A1, A4, A5, A6, B1, B2, B3, B4, B5 load on Factor 1 (GWIF: general wakefulness inability and fatigue), with a maximum possible score of 27. A2, A3 and B6 load on Factor 2 (DWIF: driving wakefulness inability and fatigue), with a maximum possible score of 9.

Table 3—Normal subjects: mean, SD, medians and percentiles

	Mean	SD	Median	Mean + 1 SD	85 th percentile	Mean + 2 SD	95 th percentile
Young adults (18-45 y)							
Age	27.4	6.4	26.0	33.8	35.0	40.2	41.6
BMI	25.6	6.0	24.4	31.6	30.9	37.6	37.6
Time in bed (h)	7.7	1.6	7.5	9.3	8.5	10.9	11.0
ESS	6.8	4.1	5.0	10.9	11.0	15.0	15.0
SWIFT	7.1	4.9	6.0	12.0	12.0	16.9	17.6
GWIF	6.6	4.4	6.0	11.0	10.0	15.4	15.6
DWIF	0.6	1.1	0.0	1.7	1.8	2.8	3.0
Middle-aged to older adults (> 45 y)							
Age	63.2	12.8	26.0	76.0	79.3	88.8	87.1
BMI	27.8	5.9	24.4	33.7	33.3	39.6	40.8
Time in bed (h)	7.8	1.4	7.5	9.2	9.0	10.6	10.1
ESS	5.8	4.0	5.0	9.8	10.7	13.8	14.6
SWIFT	4.7	4.3	6.0	9.0	9.0	13.3	13.6
GWIF	4.2	3.9	6.0	8.1	8.0	12.0	13.0
DWIF	0.5	0.9	0.0	1.4	1.7	2.3	3.0

Mean, median, standard deviation, mean + 1 and + 2 standard deviations, and 85th and 95th percentiles are shown by age group for age, BMI, time in bed, the ESS, SWIFT, and the general wakefulness inability and fatigue (GWIF), and driving wakefulness inability and fatigue (DWIF) factors of the SWIFT.

Upon performing factor analysis separately for each age group, there were the same 2 factors for young adults, accounting for the same 36% and 20% of variance. For middle-aged to older adults, Factor 2 remained the same (A2, A3, B6) and accounted for 19% of variance. Factor 1 separated into 3 factors. The new Factor 1 (A1, A4, B1, B2, B3, B5) accounted for 27% of variance, while A5 and B4 (17% of variance), and A6 (10% of variance) became new separate factors, suggesting that wakefulness inability/fatigue while reading or studying may separate from general wakefulness inability/fatigue in middle-aged to older adults.

OSA Patients

Of the 286 patients with $AHI \geq 5$, 86 were young adults (ages 18-45 years) and 200 were middle-aged to older adults (age > 45 years). The 188 normal young adults differed significantly from the 86 young adults with $AHI \geq 5$ in age, SWIFT, GWIF, DWIF, and ESS. **Table 4** gives the means and standard deviations. **Table 5** gives the AUC and, using cutoffs at greater than mean + 1 SD (> 10 for ESS, > 12 for SWIFT, > 11 for GWIF, and > 1 for DWIF), the sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and Youden's index. SWIFT and GWIF but not DWIF had better AUC, positive and negative likelihood ratios and Youden's index than ESS. **Figure 1** shows that the ROC for ESS was entirely within the ROC for SWIFT and GWIF. However, there was no significant difference between AUC for ESS and SWIFT.

The 68 normal middle-aged to older adults differed significantly from the 200 middle-aged to older adults with $AHI \geq 5$ in age, SWIFT, GWIF, DWIF, and ESS. **Table 4** gives the means and standard deviations. **Table 5** gives the AUC and, using cutoffs at greater than mean + 1 SD (> 9 for ESS, > 9 for SWIFT, > 8 for GWIF, and > 1 for DWIF), the sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and Youden's index. SWIFT, GWIF, and DWIF had better AUC, positive and negative likelihood ratios, and Youden's index than ESS. **Figure 2** shows the ROC for ESS was entirely within the ROC for SWIFT and GWIF. The AUC was significantly higher

Table 4—Normal subjects vs. OSA patients: means and SD

	Young adults (18-45 y)		Middle-aged to older adults (> 45 y)	
	Normal	With OSA (AHI \geq 5)	Normal	With OSA (AHI \geq 5)
Age*	27.4 (6.4)	37.3 (6.2)	63.2 (12.8)	58.5 (8.9)
ESS*	6.8 (4.1)	9.7 (5.5)	5.8 (4.0)	9.2 (5.5)
SWIFT*	7.1 (4.9)	12.9 (7.3)	4.7 (4.3)	12.0 (7.9)
GWIF*	6.6 (4.4)	11.4 (6.2)	4.2 (3.9)	10.4 (6.6)
DWIF*	0.6 (1.1)	1.4 (1.8)	0.5 (0.9)	1.6 (2.0)

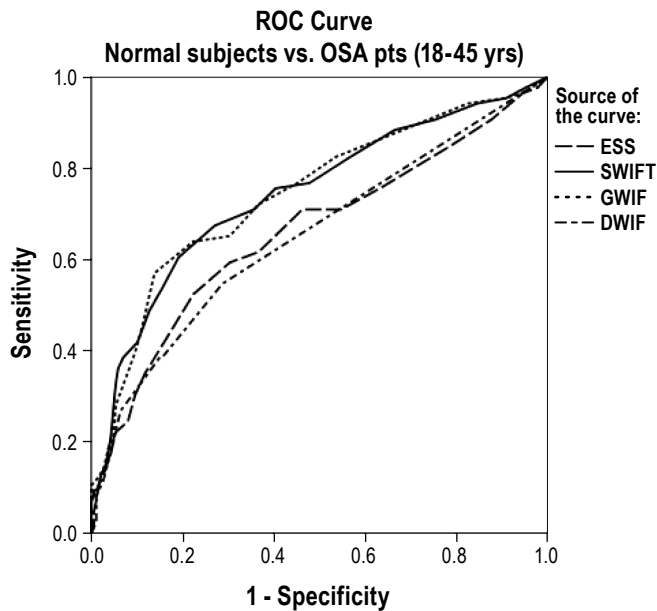
*Normal different from OSA patients at $p < 0.001$ for both age groups. All values mean (SD).

Table 5—Normal subjects vs. OSA patients: indices of test superiority

	Young adults (18-45 y)		Middle-aged to older adults (> 45 y)	
	Value	95% CI	Value	95% CI
AUC for ESS	0.660	0.585-0.734	0.688	0.620-0.757
AUC for SWIFT*	0.743	0.676-0.809	0.793	0.736-0.850
AUC for GWIF*	0.743	0.677-0.810	0.793	0.733-0.851
AUC for DWIF	0.652	0.578-0.725	0.669	0.602-0.737
sensitivity for ESS	0.453	0.368-0.535	0.430	0.395-0.457
sensitivity for SWIFT*	0.488	0.405-0.563	0.565	0.532-0.587
sensitivity for GWIF*	0.453	0.372-0.525	0.540	0.508-0.561
sensitivity for DWIF	0.384	0.302-0.462	0.430	0.397-0.453
specificity for ESS	0.819	0.780-0.857	0.809	0.706-0.887
specificity for SWIFT*	0.872	0.834-0.906	0.868	0.771-0.932
specificity for GWIF*	0.888	0.851-0.921	0.882	0.787-0.943
specificity for DWIF*	0.846	0.808-0.882	0.853	0.755-0.921
ρ_+ for ESS	2.508	1.670-3.731	2.249	1.364-4.093
ρ_+ for SWIFT*	3.826	2.445-6.016	4.269	2.351-8.794
ρ_+ for GWIF*	4.060	2.503-6.654	4.590	2.419-9.967
ρ_+ for DWIF	2.488	1.573-3.910	2.924	1.638-5.848
ρ_- for ESS	0.667	0.543-0.811	0.705	0.613-0.856
ρ_- for SWIFT*	0.587	0.482-0.713	0.501	0.443-0.607
ρ_- for GWIF*	0.615	0.516-0.737	0.525	0.466-0.625
ρ_- for DWIF	0.729	0.610-0.864	0.668	0.593-0.800
γ for ESS	0.273	0.147-0.392	0.239	0.102-0.343
γ for SWIFT*	0.361	0.239-0.469	0.433	0.303-0.519
γ for GWIF*	0.342	0.224-0.446	0.422	0.295-0.503
γ for DWIF	0.229	0.110-0.344	0.283	0.151-0.375

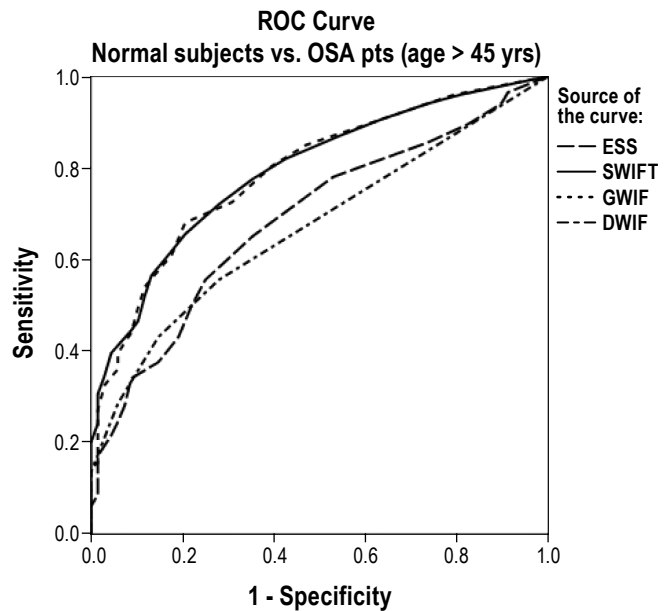
*Value superior to value for ESS in both age groups. [†]AUC significantly higher than that for ESS in middle-aged to older adults. ρ_+ and ρ_- are likelihood ratios for positive and negative test. γ , Youden's index. Mean + 1 SD used as cut-offs for calculation of sensitivity, specificity, ρ_+ , ρ_- and γ . Young adults: ESS > 10, SWIFT > 12, GWIF > 11, DWIF > 1. Middle-aged to older adults: ESS > 9, SWIFT > 9, GWIF > 8, DWIF > 1.

Figure 1—ROC curves for ESS, SWIFT, GWIF, and DWIF for normal subjects vs. OSA patients in age group 18-45 years



ESS, Epworth Sleepiness Scale; SWIFT, Sleepiness-Wakefulness Inability and Fatigue Test; GWIF, general wakefulness inability and fatigue factor; DWIF, driving wakefulness inability and fatigue factor.

Figure 2—ROC curves for ESS, SWIFT, GWIF, and DWIF for normal subjects vs. OSA patients in age group > 45 years



ESS, Epworth Sleepiness Scale; SWIFT, Sleepiness-Wakefulness Inability and Fatigue Test; GWIF, general wakefulness inability and fatigue factor; DWIF, driving wakefulness inability and fatigue factor.

for SWIFT ($z = 2.36$, $p = 0.018$) than for ESS, and the AUC was also significantly higher for GWIF than for ESS ($z = 2.35$, $p = 0.019$), but not for DWIF.

There were no significant correlations found in the OSA patients between ESS, SWIFT, GWIF, or DWIF on the one hand, and sleep efficiency, PLMAI, or lowest oxygen saturation on the other. SWIFT ($r = 0.16$, $p = 0.006$), GWIF ($r = 0.15$, $p = 0.009$) and DWIF ($r = 0.14$, $p = 0.023$), but not ESS, were significantly correlated with arousal index. ESS ($r = 0.14$, $p = 0.018$) and GWIF ($r = 0.14$, $p = 0.022$), but not SWIFT or DWIF, were significantly correlated with AHI.

CPAP Treatment

ESS, SWIFT, GWIF, and DWIF improved significantly in patients on CPAP in both age groups (36 young adults: $t = 7.1$, $df = 35$, $p < 0.001$ for ESS, $t = 7.0$, $df = 35$, $p < 0.001$ for SWIFT, $t = 7.4$, $df = 35$, $p < 0.001$ for GWIF, $t = 3.4$, $df = 35$, $p = 0.002$ for DWIF; 101 middle-aged to older adults: $t = 9.7$, $df = 100$, $p < 0.001$ for ESS, $t = 12.2$, $df = 100$, $p < 0.001$ for SWIFT, $t = 11.5$, $df = 100$, $p < 0.001$ for DWIF, $t = 7.7$, $df = 100$, $p < 0.001$ for DWIF). Effect sizes and 95% confidence intervals for the 137 subjects were as follows: ESS 0.96 (0.07, 1.63), SWIFT 1.07 (-0.20, 1.98), GWIF 1.04 (-0.03, 1.82), DWIF 0.75 (0.43, 0.93). One hundred fourteen of 137 (83.2%) subjects were compliant (use ≥ 4 h/night) for $\geq 70\%$ of nights. Compliance was significantly correlated with improvement in SWIFT ($r = 0.21$, $p = 0.015$), GWIF ($r = 0.18$, $p = 0.034$) and DWIF ($r = 0.18$, $p = 0.032$), but not ESS ($r = 0.11$, $p = 0.216$). Improvement in SWIFT ($r = 0.22$, $p = 0.011$) and GWIF ($r = 0.24$, $p = 0.004$) were also significantly correlated with AHI, but im-

provement in DWIF or ESS were not. **Table 6** gives by age group the pre- and post-treatment data, as well as numbers above and below the cutoffs before and after treatment, effect sizes, and χ^2 statistics. SWIFT, GWIF, DWIF, and ESS were all valuable in demonstrating conversion from abnormal to normal values with CPAP use.

Patients Evaluated for Narcolepsy

Of 49 patients evaluated with PSG and MSLT for evaluation of narcolepsy, 37 were young adults (ages 18-45 years), and 12 were middle-aged to older adults (age > 45 years). Ten of the young adults and none of the middle-aged to older adults met MSLT criteria for diagnosis of narcolepsy—a 20% positive diagnostic rate, which is comparable to the 20% (170 of 832) positive diagnostic rate for the MSLT reported earlier in sleepy patients without OSA.³¹ The young adults with narcolepsy were significantly younger than the young adults without narcolepsy (24.1, SD 5.3 vs. 38.4, SD 13.6), but did not significantly differ from them in ESS, SWIFT, GWIF, or DWIF. The 188 normal young adults differed significantly from the 37 young adults evaluated for suspicion of narcolepsy in SWIFT, GWIF, DWIF, and ESS, but not in age. **Table 7** gives the means and standard deviations. **Table 8** gives the AUC and, using cutoffs at greater than mean + 1 SD (> 10 for ESS, > 12 for SWIFT, > 11 for GWIF, and > 1 for DWIF), the sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and Youden's index. SWIFT, GWIF, and DWIF had better AUC, positive and negative likelihood ratios, and Youden's index than ESS. **Figure 3** shows the ROC for ESS was entirely within the ROC for SWIFT and GWIF. The AUC was significantly higher for SWIFT ($z = 2.29$, $p = 0.022$) than for

Table 6—Before and on CPAP treatment

	Mean	SD	# with high scores	# with normal scores	Effect size (95% confidence intervals)	χ^2 for high and normal scores	p-value for χ^2
Young adults (age 18-45 y)							
Before CPAP							
Sample size	36						
Age	37.56	6.20					
AHI	28.77	28.78					
ESS	10.33	5.16	18	18			
SWIFT	13.44	7.04	19	17			
GWIF	11.97	6.11	16	20			
DWIF	1.47	1.68	14	22			
On CPAP							
Compliance %	75.2	24.8					
ESS*	4.67	4.15	3	33	1.23 (-0.46,2.58)	13.2	< 0.001
SWIFT*	5.67	6.07	5	31	1.20 (-1.10,3.18)	10.6	0.001
GWIF*	5.28	5.46	5	31	1.17 (-0.83,2.95)	6.7	0.010
DWIF*	0.39	0.90	3	33	0.81 (-0.13,1.01)	7.7	0.006
Middle-aged to older adults (age > 45 y)							
Before CPAP							
Sample size	101						
Age	58.92	8.98					
AHI	47.11	3					
ESS	9.78	5.45	47	54			
SWIFT	12.83	7.74	62	39			
GWIF	11.02	6.49	61	40			
DWIF	1.81	2.00	48	53			
On CPAP							
Compliance %	81.7	22.3					
ESS*	5.64	3.93	19	82	0.88 (-0.19,1.64)	16.4	< 0.001
SWIFT*	6.13	5.22	26	75	1.02 (-0.49,2.04)	24.7	< 0.001
GWIF*	5.52	4.43	25	76	0.99 (-0.27,1.86)	24.8	< 0.001
DWIF*	0.60	1.18	16	85	0.74 (0.35-0.97)	22.0	< 0.001

*Significant improvement, p < 0.01.

Table 7—Normal subjects vs. MSLT patients: means and SD

	Young adults (18-45 y)		Middle-aged to older adults (> 45 y)	
	Normal	Evaluated for narcolepsy	Normal	Evaluated for narcolepsy
Age	27.4 (6.4)	29.2 (8.1)	63.2 (12.8)	55.0 (6.7)
ESS*	6.8 (4.1)	11.7 (5.1)	5.8 (4.0)	11.6 (6.9)
SWIFT*	7.1 (4.9)	19.8 (7.8)	4.7 (4.3)	16.8 (7.0)
GWIF*	6.6 (4.4)	16.8 (6.3)	4.2 (3.9)	14.5 (5.5)
DWIF*	0.6 (1.1)	3.0 (2.2)	0.5 (0.9)	2.3 (2.3)

*Normal different from patients evaluated for narcolepsy, significant at p < 0.05 for both age groups. All values mean (SD).

ESS. The AUC was also significantly higher for GWIF than for ESS (z = 2.07, p = 0.038), but not for DWIF.

The 68 normal middle-aged to older adults differed significantly from the 12 middle-aged to older adults evaluated for suspicion of narcolepsy in age, SWIFT, GWIF, DWIF, and ESS. **Table 7** gives the means and standard deviations. **Table 8** gives the AUC and, using cutoffs at greater than mean

+ 1 SD (> 9 for ESS, > 9 for SWIFT, > 8 for GWIF, and > 1 for DWIF), the sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and Youden’s index. SWIFT, GWIF and DWIF had better AUC, positive and negative likelihood ratios, and Youden’s index than ESS. **Figure 4** shows the ROC for ESS was entirely within the ROC for SWIFT and GWIF. However, there was no significant difference between AUC for ESS and SWIFT.

There were no significant correlations found in the patients evaluated for narcolepsy between ESS, SWIFT, GWIF, or DWIF on the one hand and sleep efficiency, AHI, PLMAI, arousal index, or number of SOREMPs on the MSLT, on the other. ESS but not SWIFT, GWIF, or DWIF was significantly negatively correlated with mean sleep latency on the MSLT (r = -0.408, p = 0.004).

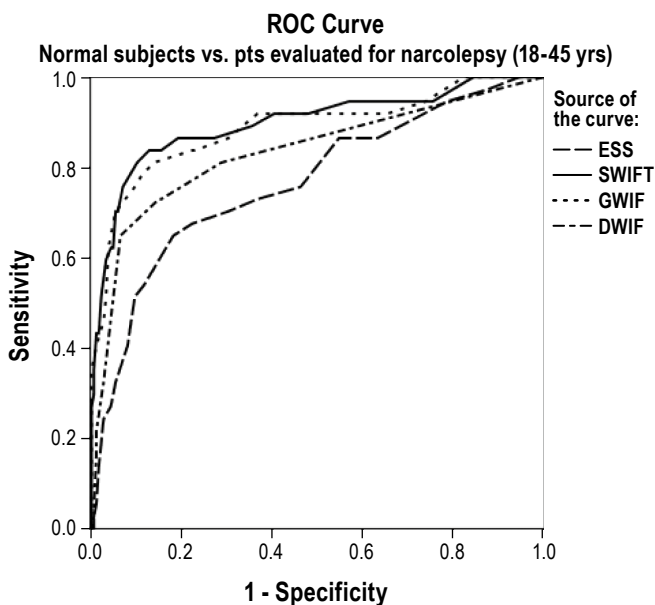
DISCUSSION

The SWIFT has high internal consistency as shown by high Cronbach α , and high test-retest reliability shown by high intra-class coefficient. Thus, the SWIFT is a reliable test.

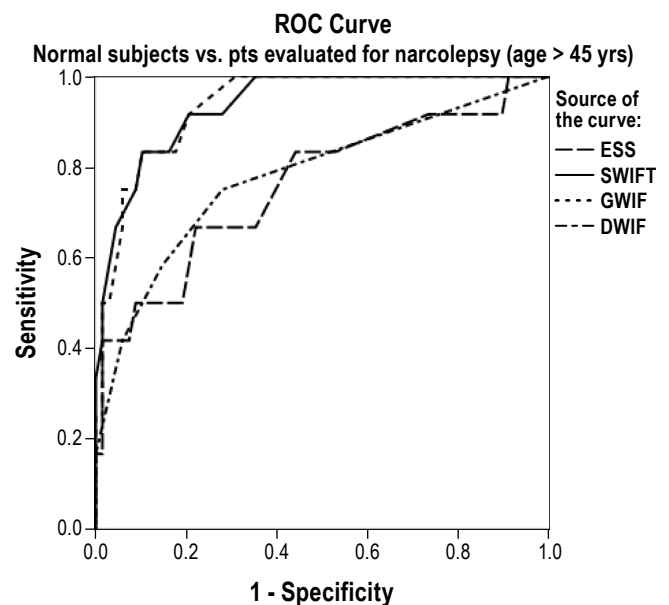
Table 8—Normal subjects vs. MSLT patients: indices of test superiority

	Young adults (18-45 y)		Middle-aged to older adults (> 45 y)	
	Value	95% CI	Value	95% CI
AUC for ESS	0.767	0.676-0.857	0.761	0.595-0.927
AUC for SWIFT [†]	0.898	0.831-0.965	0.938	0.877-0.999
AUC for GWIF [†]	0.888	0.818-0.958	0.939	0.882-0.996
AUC for DWIF [*]	0.833	0.748-0.918	0.780	0.618-0.942
sensitivity for ESS	0.649	0.494-0.781	0.500	0.237-0.758
sensitivity for SWIFT [*]	0.838	0.695-0.929	0.833	0.544-0.970
sensitivity for GWIF [*]	0.784	0.638-0.889	0.833	0.546-0.969
sensitivity for DWIF [*]	0.730	0.578-0.848	0.583	0.307-0.818
specificity for ESS	0.818	0.789-0.845	0.809	0.762-0.854
specificity for SWIFT [*]	0.872	0.844-0.890	0.868	0.817-0.892
specificity for GWIF [*]	0.888	0.860-0.909	0.882	0.832-0.906
specificity for DWIF [*]	0.846	0.816-0.869	0.853	0.804-0.894
ρ_+ for ESS	3.587	2.335-5.045	2.615	0.996-5.200
ρ_+ for SWIFT [*]	6.563	4.463-8.469	6.296	2.964-8.953
ρ_+ for GWIF [*]	7.017	4.549-9.755	7.083	3.241-10.355
ρ_+ for DWIF [*]	4.731	3.136-6.472	3.967	1.569-7.745
ρ_- for ESS	0.468	0.259-0.642	0.618	0.284-1.001
ρ_- for SWIFT [*]	0.186	0.080-0.361	0.192	0.034-0.559
ρ_- for GWIF [*]	0.243	0.123-0.421	0.189	0.034-0.546
ρ_- for DWIF [*]	0.320	0.175-0.518	0.489	0.203-0.862
γ for ESS	0.468	0.282-0.626	0.309	-0.001-0.612
γ for SWIFT [*]	0.710	0.539-0.819	0.701	0.360-0.861
γ for GWIF [*]	0.672	0.498-0.797	0.716	0.377-0.876
γ for DWIF [*]	0.575	0.393-0.717	0.436	0.111-0.712

*Value superior to value for ESS in both age groups. [†]AUC significantly higher than AUC for ESS in young adults. ρ_+ and ρ_- are likelihood ratios for positive and negative test. γ , Youden's index. Mean + 1 SD used as cut-offs for calculation of sensitivity, specificity, ρ_+ , ρ_- and γ . Young adults: ESS > 10, SWIFT > 12, GWIF > 11, DWIF > 1. Middle-aged to older adults: ESS > 9, SWIFT > 9, GWIF > 8, DWIF > 1.

Figure 3—ROC curves for ESS, SWIFT, GWIF, and DWIF for normal subjects vs. patients evaluated for narcolepsy in age group 18-45 years

ESS, Epworth Sleepiness Scale; SWIFT, Sleepiness-Wakefulness Inability and Fatigue Test; GWIF, general wakefulness inability and fatigue factor; DWIF, driving wakefulness inability and fatigue factor.

Figure 4—ROC curves for ESS, SWIFT, GWIF, and DWIF for normal subjects vs. patients evaluated for narcolepsy in age group > 45 years

ESS, Epworth Sleepiness Scale; SWIFT, Sleepiness-Wakefulness Inability and Fatigue Test; GWIF, general wakefulness inability and fatigue factor; DWIF, driving wakefulness inability and fatigue factor.

The twelve test items of the SWIFT load on to two different factors. Factor 1 seems to be a measure of general wakefulness inability and fatigue, while Factor 2 seems to measure driving wakefulness inability and fatigue, indicating that it may be possible to measure separately general and driving related concepts/symptoms.

The ability of SWIFT, GWIF, and DWIF to discriminate between normal subjects and patients with OSA, as well as patients presenting with symptoms suggesting narcolepsy, shows that the SWIFT is a valid test, as does the ability to show significant improvement with CPAP treatment of OSA.

The SWIFT and GWIF are superior to the ESS (the criterion standard) in discriminating between normal subjects and patients with OSA in both age groups, with regard to sensitivity/specificity/discriminant validity, as shown by AUC (statistically significantly so for middle-aged and older adults), Youden's index, as well as the positive and negative likelihood ratios. The SWIFT and GWIF are also superior to the ESS in discriminating between normal subjects and patients evaluated for narcolepsy in both age groups, with regard to sensitivity/specificity/discriminant validity, as shown by AUC (statistically significantly for young adults), Youden's index, and positive and negative likelihood ratios. All the ROCs for ESS (young and middle-aged to older adults, patients with OSA, and patients evaluated with narcolepsy) were entirely within the ROCs for SWIFT and GWIF. Given the rarity of narcolepsy, comparisons were made using patients evaluated for narcolepsy rather than patients diagnosed with narcolepsy. However, patients with OSA were excluded from this group; patients not positive for narcolepsy were as sleepy, fatigued, and unable to maintain wakefulness as the patients with narcolepsy. Thus, SWIFT and GWIF may be more useful than ESS in terms of clinical utility in discriminating between normal and sleep disordered subjects.

Effect sizes were similar for improvement in ESS, SWIFT, and GWIF (but lower for DWIF) with CPAP treatment in young adults. In middle-aged to older adults, effect sizes for SWIFT and GWIF were higher than those for ESS and DWIF. Comparisons of the number of patients with high ESS, SWIFT, GWIF, and DWIF before and after CPAP treatment revealed significant differences in both age groups. Improvement in SWIFT, GWIF, and DWIF, but not ESS, was significantly correlated with compliance. This compliance-response relationship lends more confidence in the use of the SWIFT or GWIF rather than the ESS is assessing treatment response with CPAP despite similar effect sizes for SWIFT, GWIF, and ESS. The finding that only ESS, but not SWIFT, GWIF, or DWIF, is correlated with MSLT suggests that subjects were able to separate the concept of tendency to fall asleep (as measured by the MSLT and the ESS) from wakefulness inability and fatigue. The finding that SWIFT, GWIF, and DWIF, but not ESS are correlated with arousal index suggests that they are a better measure of lack of sleep quality than the ESS. The finding that wakefulness inability and fatigue did not load on to separate factors on factor analysis suggests that subjects may have a hard time separating these two concepts.

The separation of data into two groups by age necessitated by a correlation between age and SWIFT as well as ESS in the combined group, provides a built-in replication, and similar findings in the two independent age groups (though more robust

in the middle-aged to older adults in the case of OSA and in young adults in the case of patients evaluated for narcolepsy) lend increased confidence to the results.

Mills et al. have reported that predictors of fatigue in OSA include BMI, depression scores, and soluble tumor necrosis factor receptor I (sTNF-RI), but not the severity of OSA as measured by AHI or mean oxygen saturation.³² Tumor necrosis factor- α (TNF- α) and interleukin-6 (IL-6) are increased in OSA and narcolepsy.³³ Adding measures of fatigue to the measurement scale for daytime functioning, and changing the measurement scale to measure wakefulness inability rather than tendency to fall asleep may improve the measurement of the daytime consequences of sleep disorders.

Masa et al.³⁴ reported habitual sleepiness affecting 3.6% of drivers, with an odds ratio of 13.7 for highway automobile accidents, and with considerable ESS overlap between these subjects and controls. 50% of habitually sleepy drivers had ESS < 9. This suggests that propensity to fall asleep in other circumstances (as measured by ESS) is neither necessary nor sufficient to cause increased risk for auto accidents. Although a sample of 10,000 subjects with 71% response rate showed no correlation between the ESS and the adverse consequence of automobile accidents, there was a correlation with dozing off while stopped in traffic.¹⁸ Increased risk for auto accidents may be the result of a complex mix of wakefulness inability, fatigue, and inattention/cognitive impairment, all of which may occur in sleep disordered or sleep deprived subjects. Measurement of increased risk for auto accidents may require questions directly related to wakefulness inability and fatigue while driving, as in the DWIF factor of the SWIFT. The question whether DWIF might be predictive of risk for auto accidents needs to be elucidated in further research.

This study was designed to determine if the SWIFT is a reliable and valid instrument, and if it is superior to the criterion standard, ESS in terms of specificity/sensitivity/discriminant ability, and therefore, possibly, clinical utility. We have shown that the SWIFT is reliable and has discriminant validity, that it has two factors (GWIF and DWIF), and that the SWIFT and its GWIF factor are superior to the ESS in discriminating between normal subjects and sleep disordered patients. These tests measure sleepiness/wake inability and assist in screening/diagnosis. However, they are not meant to discriminate between different causes of difficulties with wakefulness inability or sleepiness. SWIFT should be added to ESS in evaluating daytime consequences of sleep disorders. The two tests together comprise 20 questions and can form a quick questionnaire for use in the office to screen for sleepiness, wakefulness inability, and fatigue, with cutoffs of > 10 for ESS, > 12 for SWIFT, > 11 for GWIF, > 1 for DWIF in young adults (ages 18-45 years), and with cutoffs of > 9 for ESS, > 9 for SWIFT, > 8 for GWIF, > 1 for DWIF in middle-aged to older adults (age > 45 years).

A limitation of this study is that item selection was based on face validity rather than qualitative evaluation using patient focus groups. Another limitation is that the control group was recruited by means of a group e-mail to medical students and by personal solicitation in public places, and it is not clear whether this cohort of normal subjects generalizes to the population and whether it is comparable to the patient groups presented. Further, although the SWIFT and its factors are a better mea-

sure for differentiating between normal subjects and sleep disordered patients than the ESS, the areas under the curve still leave a lot to be desired. However, though there may eventually be a simple blood test to measure sleepiness, wakefulness inability, and fatigue, for now we are left with questionnaires as possibly the best proxies, though objective measures such as the psychomotor vigilance test or the divided attention driving test are other candidates.³⁵ This study was a clinical rather than an experimental study. Since the MWT is not routinely performed clinically, this study did not compare the SWIFT with the MWT. Future directions might include a study of the SWIFT using the MWT.

REFERENCES

- 2002 "Sleep in America" Poll. National Sleep Foundation website. <http://www.sleepfoundation.org/sites/default/files/2002SleepInAmericaPoll.pdf>. Published 2002. Accessed February 11, 2011.
- Drowsy Driving and Automobile Crashes. National Highway Safety Administration website. http://www.nhtsa.gov/people/injury/drowsy_driving1/drowsy.html. I: Introduction. Accessed May 22, 2012.
- Chervin RD. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Chest* 2000;118:372-9.
- Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 1986;9:519-24.
- Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540-5.
- Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;15:376-81.
- Arand D, Bonnet M, Hurwitz T, Mitler M, Rosa R, Sangal RB. The clinical use of the MSLT and MWT. *Sleep* 2005;28:123-44.
- Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment in patients with excessive sleepiness. *Electroencephalogr Clin Neurophysiol* 1982;53:658-61.
- Doghramji K, Mitler MM, Sangal RB, et al. A normative study of the maintenance of wakefulness test (MWT). *Electroencephalogr Clin Neurophysiol* 1997;103:554-62.
- Sangal RB, Thomas L, Mitler MM. The maintenance of wakefulness test (MWT) and the multiple sleep latency test (MSLT) measure different abilities in patients with sleep disorders. *Chest* 1992;101:898-902.
- Browman CP, Gujavarty KS, Sampson MG, Mitler MM. REM sleep episodes during the maintenance of wakefulness test in patients with sleep apnea syndrome and patients with narcolepsy. *Sleep* 1983;6:23-8.
- Sangal RB, Thomas L, Mitler MM. Disorders of excessive sleepiness: treatment improves ability to stay awake but does not reduce sleepiness. *Chest* 1992;102:699-703.
- Sanford SD, Lichstein KL, Durrence HH, Riedel BW, Taylor DJ, Bush AJ. The influence of age, gender, ethnicity, and insomnia on Epworth Sleepiness Scores: a normative U.S. population. *Sleep Med* 2006;7:319-26.
- Sangal RB, Sangal JM, Belisle C. Subjective and objective indices of sleepiness (ESS and MWT) are not equally useful in patients with sleep apnea. *Clin Electroencephalogr* 1999;30:73-5.
- Sangal RB, Mitler MM, Sangal JM. Subjective sleepiness ratings (Epworth Sleepiness Scale) do not reflect the same parameter of sleepiness as objective sleepiness (maintenance of wakefulness test) in patients with narcolepsy. *Clin Neurophysiol* 1999;110:2131-5.
- Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703-10.

- Benbadis SR, Mascha E, Perry MC, et al. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med* 1999;130:289-92.
- Gander PH, Marshall NS, Harris RB, Reid P. Sleep, sleepiness and motor vehicle accidents: a national survey. *Austr N Z J Public Health* 2005;29:16-21.
- Merriam-Webster Dictionary. <http://www.merriam-webster.com>. Accessed January 20, 2012.
- Sangal RB. When is sleepiness a disease? How do we measure it? *Sleep Med* 2006;7:310-1.
- Stein KD, Martin SC, Hann DM, Jacobsen PB. A multidimensional measure of fatigue for use with cancer patients. *Cancer Pract* 1998;6:143-52.
- Stein, KD, Jacobsen PB, Blanchard CM, Thors CT. Further validation of the Multidimensional Fatigue Symptom Inventory-Short Form (MFSI-SF). *J Pain Symptom Manage* 2004;27:14-23.
- Curran-Everett D. Multiple comparisons: philosophies and illustrations. *J Physiol Regul Integr Comp Physiol* 2000;279:R1-R8.
- Iber C, Ancoli-Israel S, Chesson AL, Quan SF. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine, Westchester, IL, 2007.
- Bewick V, Cheek L, Ball J. Statistics review 13: Receiver operating characteristic curves. *Crit Care* 2004;8:508-12.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
- Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley & Sons, 1981:sec 5.6.
- Hilden J, Glasziou P. Regret graphs, diagnostic uncertainty and Youden's Index. *Stat Med* 1996;15:969-86.
- Biggerstaff BJ. Comparing diagnostic tests: a simple graphic using likelihood ratios. *Stat Med* 2000;19:649-63.
- Aldrich MS, Chervin RD, Malow BA. Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy. *Sleep* 1997;20:620-9.
- Vgontzas AN, Papanicalou DA, Bixler EO, Kales A, Tysom K, Chrousos GP. Elevation of plasma cytokines in disorders of excessive daytime sleepiness: role of sleep disturbance and obesity. *J Clin Endocrinol Metab* 1997;82:1313-6.
- Mills PJ, Kim J-H, Bardwell W, Hong S, Dimsdale JE. Predictors of fatigue in obstructive sleep apnea. *Sleep Breath* 2008;12:397-9.
- Masa JF, Rubio M, Findley LJ. Habitually sleepy drivers have a higher frequency of automobile crashes associated with respiratory disorders during sleep. *Am J Respir Crit Care Med* 2000;162:1407-12.
- Sunwoo BY, Jackson N, Maislin G, Gurubhagavatula I, Goerge CF, Pack AI. Reliability of a single objective measure in assessing sleepiness. *Sleep* 2012;35:149-58.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication October, 2011

Submitted in final revised form May, 2012

Accepted for publication May, 2012

Address correspondence to: R. Bart Sangal, M.D., Professor, Oakland University William Beaumont School of Medicine, Director, Sleep Disorders Institute, 44344 Dequindre Rd. Ste 360, Sterling Heights, MI 48314; Tel: (586) 254-0707; Fax: (586) 254-7207; E-mail: sangalrb@sleepwebmd.com

DISCLOSURE STATEMENT

This was not an industry supported study. The author has indicated no financial conflicts of interest.