https://doi.org/10.5664/jcsm.8350

JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Low repeatability of Epworth Sleepiness Scale after short intervals in a sleep clinic population

Fabian A. Grewe, MD[1]; Maurice Roeder, MD[1]; Matteo Bradicich, MD[1]; Esther I. Schwarz, MD[1,2]; Ulrike Held, PhD[3]; Sira Thiel, MD[1]; Thomas Gaisl, MD[1]; Noriane A. Sievi, MD[1]; Malcolm Kohler, MD[1,2]

[1]Department of Pulmonology, University Hospital Zurich, Zurich, Switzerland; [2]Centre of Competence Sleep and Health Zurich, University of Zurich, Zurich, Switzerland; [3]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Study Objectives: The purpose of this study was to evaluate the short-term repeatability of the Epworth Sleepiness Scale (ESS) in patients with suspected obstructive sleep apnea and to determine whether transitory sleepiness of the patient influenced ESS results.

Methods: Adult participants with suspected obstructive sleep apnea taking part in a study on the diagnostic accuracy of repeated sleep studies were eligible. For assessment of repeatability, the agreement between 2 sequential ESS scores obtained within 1 day (same-day group) or on different days within 1 week (same-week group) was evaluated. By analyzing the within-day repeatability, a possible influence of situational sleepiness on ESS results was assessed. By comparing correlations of sequential scores between both groups, a potential influence of test day–specific sleepiness on ESS results was evaluated. Data were analyzed using Bland-Altman plots, intraclass correlation coefficients, standard error of measurement analysis, and relative amounts of ESS discrepancies beyond 2, 3, 5, and 7 points.

Results: Forty participants (mean age, 47.7 ± 15.4 years; 67.5% men) were included in this study, with 20 in each group. Bland-Altman analysis demonstrated considerable variability of repeated scores (mean ± 1.96 × SD = 1.93 [−3.81 to 7.66]). Discrepancies of at least 3 points between sequential ESS scores were found in 48% of all participants. Comparison of ESS repeatability between both groups showed no evidence for a difference.

Conclusions: A clinically relevant variability in ESS scores was found, even when repeated on the same day, possibly because of situational sleepiness influencing ESS results. Changes in ESS in response to interventions should be interpreted with caution because of its low test-retest reliability.

Keywords: daytime sleepiness, Epworth Sleepiness Scale, test-retest reliability, repeatability

Citation: Grewe FA, Roeder M, Bradicich M, et al. Low repeatability of Epworth Sleepiness Scale after short intervals in a sleep clinic population. *J Clin Sleep Med*. 2020;16(5):757–764.

BRIEF SUMMARY

Current Knowledge/Study Rationale: The availability of a reliable and quick test is of central importance for the diagnosis and treatment of sleep-disordered breathing. The Epworth Sleepiness Scale is the most commonly used test for assessing daytime sleepiness. However, the questionnaire's reliability is currently under discussion, although it has only been evaluated in test-retest settings using retest intervals of more than 2 months in sleep clinic populations thus far.

Study Impact: Therefore, the variability of sequential tests observed in these studies might be explained by true changes in average daytime sleepiness. We assessed the test-retest reliability of sequential tests within 1 week, aiming for a significant evaluation of the questionnaire's reliability.

INTRODUCTION

Excessive daytime sleepiness, with a prevalence of approximately 18%[1] in the general population, is associated with an increased risk for car accidents, diabetes mellitus, cardiovascular disease, and overall mortality.[2] The Epworth Sleepiness Scale (ESS) is the most widely used test for assessing levels of self-reported daytime sleepiness.[3] A change in ESS over time is used in clinical practice to assess the effect of interventions and serves as an endpoint in clinical trials.[4,5] The ESS is also implemented into treatment recommendations for patients with sleep-disordered breathing, and in some countries, ESS scores even influence prioritization for sleep investigation. However, concern about the questionnaire's repeatability, and therefore its reliability, is mounting.[6,7]

The reliability of a scale is preferably determined by comparing consecutive test scores of the same individual, obtained under comparable conditions, without interventions in between.[8–10] Although ESS validation studies, limited to healthy participants or mixed populations, found reliability to be moderate to good,[11–14] 2 studies investigating patients with suspected obstructive sleep apnea (OSA) indicated poor reliability of the ESS.[6,15] However, the time interval between test and retest was more than 2 months in the latter 2 studies. A maximum of 4 weeks is generally recommended to prevent measuring true changes in average sleep propensity.[16,17] Although intended to measure the long-term "average sleep propensity in daily life,"[18] "as distinct from feelings of sleepiness at a particular time,"[19] Slater et al.[20] assumed that transitory factors, such as sleep quantity and quality during the

previous night, might influence ESS results, thus causing variable scores depending on the situation or the test day.

Our aim was to investigate whether ESS scores remained stable within short retest intervals (1–6 days) in patients with suspected OSA. Furthermore, we evaluated whether situational sleepiness or test day–specific levels of sleepiness influenced ESS results by comparing 2 ESS scores obtained within the same day and on different days within 1 week.

METHODS

Participants

This was a subinvestigation of an ongoing, prospective sleep-cohort study assessing the effect of repeated sleep studies on the diagnostic accuracy in patients with suspected OSA, referred to a tertiary sleep center (NCT03819361). Adult patients undergoing an in-hospital sleep study within the course of a comprehensive sleep evaluation were eligible if they participated in the aforementioned trial between February and July 2019. Exclusion criteria were previous OSA diagnosis or continuous positive airway pressure (CPAP) therapy, acutely life-threatening illness, psychological constraint, and pregnancy. Participants were included in this subanalysis consecutively if they filled in the ESS questionnaire 2 times within the same day (same-day group) or on different days, with an interval of 1–6 days between (same-week group). The questionnaires had to be dated on top; thus, the date of completion could be determined exactly. All questionnaires were completed in the absence of a physician. We did not inform the participants about our intention to study the ESS repeatability. The ethics commission of Zurich approved this data analysis with BASEC-NR 2018-02305.

Measurements

The ESS is an 8-item Likert-based questionnaire. The participant is asked to estimate the propensity to doze off in 8 different situations, thereby referring to everyday life during the previous few weeks to few months.[21] Each of these situations can be rated from 0 to 3: 0 = "would never doze," 1 = "slight chance of dozing," 2 = "moderate chance of dozing," 3 = "high chance of dozing." Total score ranges from 0 to 24 points.[19] A result of 11 points or more is considered to represent pathologic daytime sleepiness.[19,22] We used the validated German version of the ESS.[23] The minimal clinically important difference (MCID) between ESS scores in response to OSA treatment is reported to be 2 points in onestudy[24] and between 2 and 3 points in another study.[25]

The following data were obtained: age, sex, body mass index (BMI), Mallampati score (range, 1–4), tongue size (range, 1–4), tonsil size (range, 1–4), alcohol consumption (yes/no), sedative medication (yes/no), apnea-hypopnea index (AHI), and the respective ESS scores. AHI values were acquired during a full-night in-hospital respiratory polygraphy or polysomnography (Alice 6 System; Respironics, Pittsburgh, Pennsylvania). Sleep studies were scored by sleep specialists according to current guidelines.[26]

Sample size calculation

Sample size calculation was performed for intraclass correlation coefficient (ICC) calculation in a 1-way random effects model, resulting in n = 20 participants per group with 2 observations per participant, for a power of 90%, $\alpha = 0.05$, and an estimated ICC[11–14] of at least 0.6.[27]

Repeatability

The repeatability of a questionnaire is concerned with the degree to which repeated measurements in stable persons under comparable conditions provide similar answers, and it is assessed in a test-retest setup.[28] Differences between the first and second ESS score were calculated for each participant. Bland-Altman analysis using the mean difference and the standard error of the mean difference was used for assessing the agreement between both scores. 95% limits of agreement (mean difference ± 1.96 × SD of the difference) were calculated as an estimate of repeatability.[29,30] Additionally, the proportion of participants showing an ESS change of more than 2, 3, 5, and 7 points between test and retest was calculated. The standard error of measurement (SEM) of the ESS was computed for both groups separately and together, using the formula: $\text{SEM} = \frac{\text{SD Difference}}{\sqrt{2}}$. The SEM signifies fluctuations in measurement results around a participant's true value, and it is a critical component of test-retest reliability evaluation.[28]

Influence of transitory sleepiness on ESS scores

ICCs were calculated to assess the correlation between the first and second ESS scores in both respective groups. By analyzing ICC and Bland-Altman plots in the same-day group, differences between ESS scores within the same day were investigated. Because the ESS measures daytime sleepiness, significant differences between ESS scores obtained from the same participant within the same day might represent an impact of situational sleepiness on questionnaire results. For evaluating the influence of test day–specific sleepiness on ESS scores, ICC of both groups were compared.

Statistical analysis

Descriptive statistics include mean and SD for continuous parameters, as well as median and quartiles for nonnormal variables. Categorical variables are shown as numbers and percentages of total. Univariate regression analysis was performed for evaluating the association between age, sex, AHI, baseline ESS scores, BMI, and the difference between sequential ESS scores of all participants. Furthermore, univariate regression analysis was conducted to investigate associations between alcohol consumption or sleep medication intake and differences between repeated ESS tests within the same day. We used STATA/SE15.1 (StataCorp, College Station, Texas) for analysis.

RESULTS

Participants

We investigated data of 47 participants in total, 7 of which dropped out because the ESS dating could not be identified precisely. Forty participants (mean age, 47.7 ± 15.4 years;

Table 1—Baseline characteristics.

Clinical Characteristics	Same Day (n = 20)	Same Week[a] (n = 20)
Age (years)	50.7 (14.50)	44.75 (16.04)
Sex, male/female (% male)	14/6 (70%)	13/7 (65%)
BMI (kg/m^2)	28.9 (26.12/33.53)	30.4 (22.28/33.90)
Mallampati score (n/4)	2 (1.5/3.5)	2 (1/3)
Tongue size (n/4)	2 (2/3)	2 (2/3)
Tonsil size (n/4)	1 (1/1.5)	1 (1/1)
Alcohol[b] (yes/no)	7/13	8/11[c]
Sedating medication[d] (yes/no)	14/6	15/4[c]
AHI (events/h)	14.1 (5.9/23.4)	6.9 (2.4/23.3)
ESS–first test (n/24)	7.4 (4.37)	8.5 (4.88)
ESS–second test (n/24)	9.9 (4.47)	9.9 (5.21)
ESS–difference[e] ($n_{(\text{second test} - \text{first test})}$)	+2.45 (3.03)	+1.4 (2.87)

Values are presented as mean (SD) or median (quartiles), unless otherwise stated. AHI = apnea-hypopnea index, BMI, body mass index, ESS = Epworth Sleepiness Scale. [a]Days between test and retest: 1 day (n = 13); 2 days (n = 4); 4 days (n = 2); 6 days (n = 1). [b]Timing of alcohol intake: evening/night (77%); with meals (23%). [c]n = 19. [d]Timing of sleep medication intake: before going to bed (90%); unknown (10%). [e]Differences between the first and the second ESS were normally distributed.

67.5% men) were included in the analysis: 20 in the same-day and 20 in the same-week groups. Baseline characteristics of both groups are listed in **Table 1**.

Short-term repeatability of the ESS

Bland-Altman analysis (**Figure 1**) demonstrated considerable variability of ESS scores in both the same-day group (mean ± 1.96 × SD = 2.45 [−3.35 to 8.25]) and the same-week group (mean ± 1.96 × SD = 1.40 [−4.09 to 6.89]), as well as in the whole study population (mean ± 1.96 × SD = 1.93 [−3.81 to 7.66]). Computation of the SEM showed similar errors in the total cohort (2.09 points), the same-day (2.14 points), and the same-week (2.03 points) groups. Discrepancies of at least 2 points between sequential ESS scores occurred in 63%, at least 3 points in 48%, at least 5 points in 20%, and at least 7 points in 8% of the total of 40 participants. **Table 2** shows differences between sequential ESS scores in both groups and the whole study cohort along results of previously published studies.

Influence of transitory sleepiness on ESS scores

The ICC was 0.65 (95% confidence interval [CI]: 0.31–0.84) in the same-day and 0.81 (95% CI: 0.58–0.92) in the same-week group, demonstrating low within-day repeatability and thus a possible influence of situational sleepiness on ESS scores. There was no evidence for a significant difference in ICC values between the groups; thus, no significant impact of test day-specific sleepiness on ESS repeatability was seen. **Figure 2** shows mean ICC values with 95% confidence intervals.

Influence of baseline variables on ESS repeatability

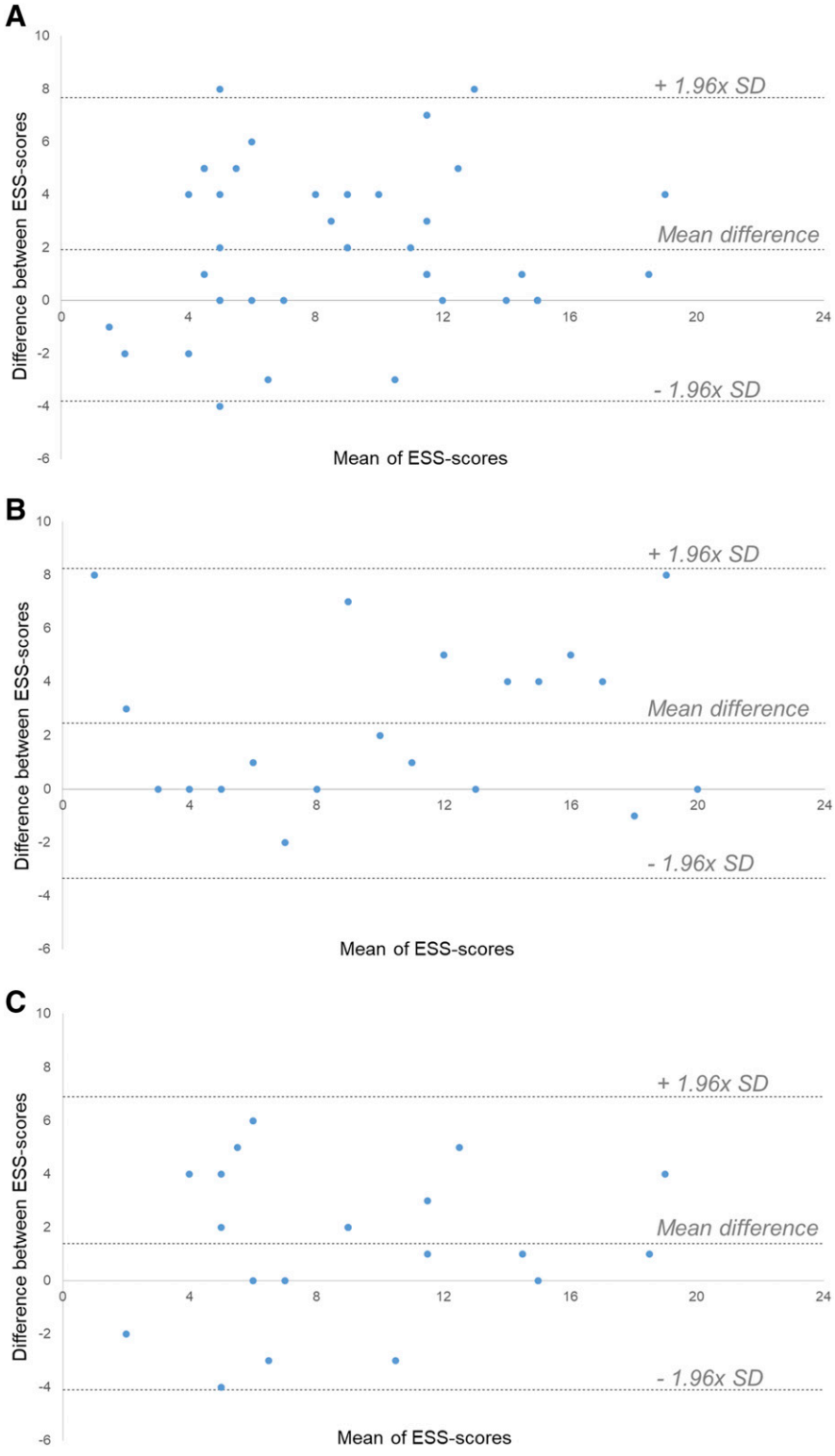
Age, sex, AHI, baseline ESS scores, and BMI were not significantly associated with differences between repeated ESS scores of all participants taken together. Furthermore, alcohol consumption and sleep medication intake were not significantly associated with differences between ESS scores obtained within the same day (**Table 3**).

DISCUSSION

To our knowledge, this is the first study to investigate the repeatability of ESS scores within time intervals as short as 1 day or 1 week. Our study found insufficient test-retest reliability even when retesting within 1 day. There was no evidence for an influence of test day–specific sleepiness on ESS scores within a maximal interval of 1 week; however, situational sleepiness might influence ESS scores, as hinted by the low within-day repeatability. These results suggest that varying levels of sleepiness in different test situations cause clinically relevant fluctuations of sequential ESS scores in patients with suspected OSA.

Quantifying sleepiness is essential for making treatment recommendations in patients with sleep-disordered breathing and for assessing treatment effects of interventions aimed to reduce sleepiness in clinical practice and in research. The ESS is cost-effective and quick, and it is the most widely used method to evaluate self-reported daytime sleepiness. Nevertheless, validation of the questionnaire has been limited to comparisons with vague indicators of sleepiness, such as sleep disorder severity, because the true average sleep propensity cannot be quantified with certainty. Objective measurement of sleepiness can be conducted by means of multiple sleep latency testing (MSLT); however, MSLT determines situational sleep propensity, whereas the ESS is intended to measure long-term average sleep propensity.[3] Some studies found no

Figure 1—Mean versus difference of sequential ESS scores.



Bland-Altman plots for **(A)** all participants, **(B)** the same-day group, and **(C)** the same-week group. The difference between 2 consecutive ESS scores is plotted against their mean for analyzing levels of accordance between 2 measurements. The blue dots depict individual measurements.

association between ESS and MSLT,[31,32] whereas others showed a correlation[33,34] even though the methodology of some studies remains disputable.

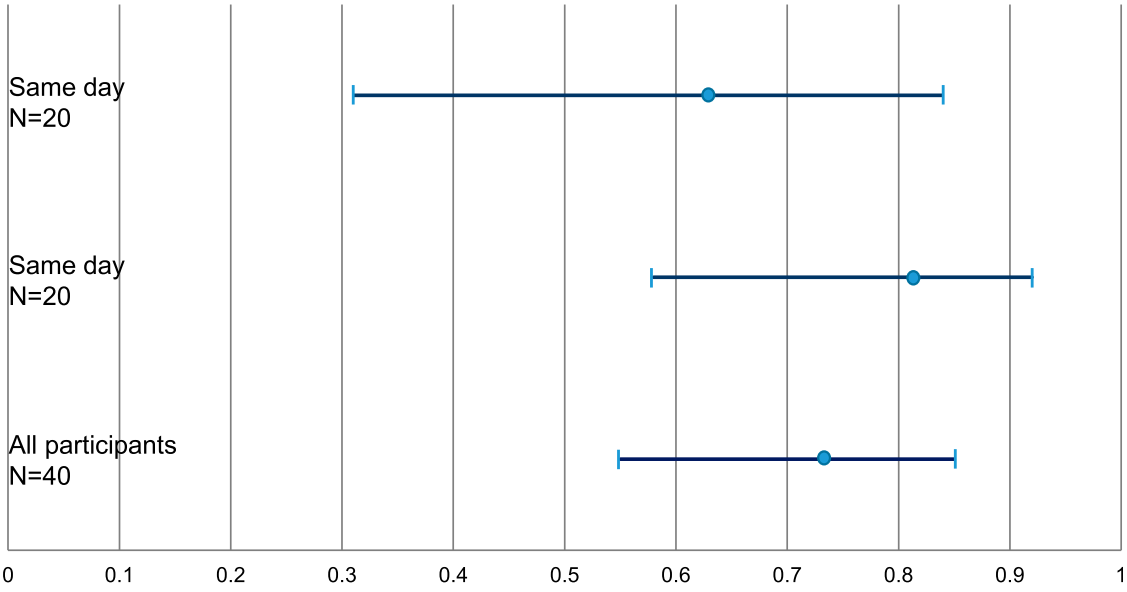
Because comparing ESS scores with other measures is not applicable, the questionnaire's reliability is a topic of utmost relevance. Reliability is a crucial psychometric property of

Table 2—Relative values of participants showing differences of at least 2, 3, 5, and 7 points between sequential ESS scores in the current and in previous studies.

	Bloch[23]	Johns[14]	Chung[13]	Nguyen[15]	Campbell[6]	Current study–all participants	Current study–same day	Current study–same week
Interval	5 months	5 months	3 months	71 (92) days	<6 months	0.5 (0/1) days	0 days	1 (1/2) days
N	19	87	56	142	154	40	20	20
Correlation coefficient	NA	Pearson: 0.822	Spearman: 0.72	Pearson: 0.73	Pearson: 0.45	ICC: 0.73	ICC: 0.65	ICC: 0.81
Population	Hospital employees	Medical students	Mixed	Patients with suspected OSA				
ESS difference ≥ 2	NA	48%	46%	61%	61%	63%	55%	70%
ESS difference ≥ 3	26%	18%	27%	41%	46%	48%	45%	50%
ESS difference ≥ 5	NA	3%	4%	23%	21%	20%	25%	15%
ESS difference ≥ 7	NA	NA	NA	10%	8%	8%	15%	0%

Values are presented as percentage of the respective study participants, as mean (SD) or as median (quartiles), unless otherwise stated. AHI = apnea-hypopnea index, ESS = Epworth Sleepiness Scale, ICC = intraclass correlation coefficient, Interval = time interval between test and retest, n = number of study participants, NA, not available.

Figure 2—ICC mean values with 95% confidence intervals.



ICCs (blue dots) with the corresponding 95% confidence intervals (blue vertical bar) of the same-day group, the same-week group, and the whole cohort. No significant difference between the 3 depicted coefficients is evident, as all confidence intervals overlap considerably.

a scale, preferably investigated in a test-retest setup.[8–10] Validation studies limited to healthy participants or mixed populations found the ESS test-retest reliability to be adequate.[11–14] Only 1 validation study calculated the SEM, and all of the aforementioned studies focused on correlation coefficients. We performed Bland-Altman analysis, which is superior to correlation coefficients in reflecting the agreement between 2 measurements.[29] The resulting plots demonstrate considerable variability of consecutive ESS scores even when testing repetitively during the same day. If repeated tests without interventions between show significant fluctuations, these fluctuations must be taken into account when using the ESS for quantifying effects of sleepiness-reducing interventions. Two meta-analyses of randomized controlled trials on the impact of CPAP therapy on daytime sleepiness show treatment effects of −2.43 points (95% CI, −1.92 to −2.95), and of −2.5 points (95% CI, 2.0–2.9), respectively.[35,36] The consistent reduction of ESS scores in response to CPAP therapy observed in the aforementioned meta-analyses shows that the ESS measures daytime sleepiness as it is intended to. However, the SEM found in our study was more than 2 points in both groups, suggesting that the amount of ESS change in response to CPAP treatment described in clinical trials might be significantly influenced by measurement error. Furthermore, considering

Table 3—Linear regression analysis.

Parameter	Coefficient	95% CI	*P* value
Univariate linear regression of the influence of baseline parameters on the difference between second and first ESS scores			
Age	0.05	0 to 0.09	.051
Male sex	1.94	−0.01 to 3.89	.051
AHI	−0.05	−0.11 to 0.00	.054
Baseline ESS scores	0.01	−0.01 to 0.03	.234
BMI	−0.05	−0.16 to 0.05	.312
Univariate linear regression of the influence of alcohol consumption and sleep medication intake on ESS differences within the same day			
Alcohol	1.57	−1.40 to 4.54	.281
Sleep medication	1.02	−2.32 to 4.18	.504

AHI = apnea-hypopnea index, BMI = body mass index, CI = confidence interval, ESS = Epworth Sleepiness Scale.

the SEM being almost equal to the MCID (which is 2−3 points), the ESS might not be accurate in confining clinically relevant effects of OSA treatment. Our study is in line with the results of Nguyen et al[15] and Campbell et al,[6] who also showed that ESS discrepancies between repeated measures exceed the MCID. Differences of at least 3 points between repeated ESS scores occurring in 41–50% reinforce concerns about the longitudinal application of the questionnaire.[6,15] Aforementioned validation studies[11–14] found greater agreement between sequential ESS scores. Participants of the studies of Nguyen's et al[15] and Campbell et al[6] and this study were all patients with suspected OSA; therefore, ESS scores might show greater variability with higher total scores.

Slater et al[20] supposed that transitory levels of sleepiness might influence performance in the ESS. If the participant is sleepier on the specific day or in the specific situation of ESS testing, he or she might state a higher propensity to fall asleep in a certain situation from his or her current point of view. If test day−specific sleepiness influenced ESS results, 2 scores obtained during the same day should show higher levels of agreement than 2 scores obtained on different days. We therefore compared the correlation of 2 consecutive ESS scores obtained within 1 day with the correlation of 2 scores obtained during different days. There was no evidence for a difference between the 2 ICCs or between the 1.96 × SD ranges of related Bland-Altman plots. Therefore, test day−specific sleepiness might not be the reason for the low repeatability of ESS scores. Objective methods for sleepiness assessment, such as the MSLT, the Maintenance of Wakefulness Test, or the Oxford Sleep Resistance test, show variable results within 1 day, as they all measure situational sleepiness.[3,37] As the ICC and the Bland-Altman plot showed considerable variability of scores in the same-day group, within-day repeatability of ESS results appears to be quite poor. Thus, ESS results might be influenced by the participant's sleepiness during the specific test situation rather than by sleepiness specific for the test day.

One limitation of our study might be the short retest interval. There is no general rule determining a minimum retest interval. Nonetheless, most statisticians conclude that 1 or 2 weeks should be the smallest period; otherwise, memory effects could cause greater agreement between consecutive tests.[28] However, repeatability of the ESS was even lower in the same-day group compared with the same-week group. Furthermore, retesting on the same day was on purpose for assessing the influence of transitory sleepiness on ESS scores. Another limitation is that the power of our study was limited for linear regression analysis. Thus, further studies will be necessary to investigate potential associations between patient characteristics and differences between repeated ESS scores.

CONCLUSIONS

We found a clinically relevant variability of the ESS even when testing repeatedly on the same day. There was no evidence for an impact of test day−specific sleepiness on ESS repeatability. However, the within-day variability of scores suggests that situational sleepiness might influence ESS results, and thus might be the root cause for the considerable SEM. This study suggests that the reliability of the ESS is not adequate to provide the basis for clinical decisions or to assess treatment effects, because the baseline fluctuation of scores reaches or exceeds the MCID. Therefore, our findings call into question the current fashion of use of the questionnaire in research and in clinical settings. Comparing 2 ESS scores from the same participant might not be applicable for longitudinal monitoring of sleep disorders and for assessing the effectiveness of interventions aimed to reduce sleepiness.

As an implication for future research, we propose the investigation of an average value of repeated ESS scores. For instance, by combining the results of 2 sequential ESS questionnaires, the reliability might be increased and therefore provide the basis for determining effects of sleepiness-reducing interventions in research and in clinical settings. Furthermore, validated questionnaires for sleepiness testing besides the ESS, such as the Stanford Sleepiness Scale or the Sleep-Related Impairment domain of the Patient-

Reported Outcomes Measurement Information System, are less commonly used. Future studies are needed to address the repeatability of these questionnaires to find alternative methods of sleepiness testing with possibly superior reliability.

ABBREVIATIONS

AHI, apnea hypopnea index
BMI, body mass index
CI, confidence interval
CPAP, continuous positive airway pressure
ESS, Epworth Sleepiness Scale
ICC, intraclass correlation coefficient
MCID, minimal clinically important difference
MSLT, multiple sleep latency test
OSA, obstructive sleep apnea
SEM, standard error of measurement

REFERENCES

1. Swanson LM, Arnedt JT, Rosekind MR, Belenky G, Balkin TJ, Drake C. Sleep disorders and work performance: findings from the 2008 National Sleep Foundation Sleep in America poll. *J Sleep Res*. 2011;20(3):487–494.
2. Endeshaw Y, Rice TB, Schwartz AV, et al. Snoring, daytime sleepiness, and incident cardiovascular disease in the health, aging, and body composition study. *Sleep*. 2013;36(11):1737–1745.
3. Kendzerska TB, Smith PM, Brignardello-Petersen R, Leung RS, Tomlinson GA. Evaluation of the measurement properties of the Epworth sleepiness scale: a systematic review. *Sleep Med Rev*. 2014; 18(4):321–331.
4. Antic NA, Catcheside P, Buchan C, et al. The effect of CPAP in normalizing daytime sleepiness, quality of life, and neurocognitive function in patients with moderate to severe OSA. *Sleep*. 2011;34(1):111–119.
5. Hardinge FM, Pitson DJ, Stradling JR. Use of the Epworth Sleepiness Scale to demonstrate response to treatment with nasal continuous positive airways pressure in patients with obstructive sleep apnoea. *Respir Med*. 1995;89(9):617–620.
6. Campbell AJ, Neill AM, Scott DAR. Clinical reproducibility of the Epworth Sleepiness Scale for patients with suspected sleep apnea. *J Clin Sleep Med*. 2018;14(5):791–795.
7. Quan SF. Abuse of the Epworth Sleepiness Scale. *J Clin Sleep Med*. 2013;9(10):987.
8. Polit DF, Yang FM. *Measurement and the Measurement of Change: A Primer for Health Professionals.* Wolters Kluwer, Philadelphia; 2014.
9. de Vet HCWT, Caroline B. Reliability. In: Mokkink LB, Knol DL, eds. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011:96-149.
10. U.S Department of Health and Human Services FDA Center for Dug Evaluation and Research, FDA Center for Biologics Evaluation and Research, FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4(1):79.
11. Bourke SC, McColl E, Shaw PJ, Gibson GJ. Validation of quality of life instruments in ALS. *Amyotroph Lateral Scler Other Motor Neuron Disord*. 2004;5(1):55–60.
12. Cho YW, Lee JH, Son HK, Lee SH, Shin C, Johns MW. The reliability and validity of the Korean version of the Epworth sleepiness scale. *Sleep Breath*. 2011;15(3):377–384.
13. Chung KF. Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosom Res*. 2000;49(5):367–372.
14. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*. 1992;15(4):376–381.
15. Nguyen AT, Baltzan MA, Small D, Wolkove N, Guillon S, Palayew M. Clinical reproducibility of the Epworth Sleepiness Scale. *J Clin Sleep Med*. 2006;2(2):170–174.
16. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, UK: Oxford University Press; 2015.
17. Price L. *Psychometric Methods: Theory into Practice, December 2016*. New York: Guilford Publications; 2016.
18. Johns MW. A new perspective on sleepiness. *Sleep Biol Rhythms*. 2010;8(3):170–179.
19. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14(6):540–545.
20. Slater G, Steier J. Excessive daytime sleepiness in sleep disorders. *J Thorac Dis*. 2012;4(6):608–616.
21. Johns MW. Epworth Sleepiness Scale. https://epworthsleepinessscale.com/about-the-ess/. Accessed September 13, 2019.
22. Onen F, Moreau T, Gooneratne NS, Petit C, Falissard B, Onen SH. Limits of the Epworth Sleepiness Scale in older adults. *Sleep Breath*. 2013;17(1):343–350.
23. Bloch KE, Schoch OD, Zhang JN, Russi EW. German version of the Epworth Sleepiness Scale. *Respiration*. 1999;66(5):440–447.
24. Crook S, Sievi NA, Bloch KE, et al. Minimum important difference of the Epworth Sleepiness Scale in obstructive sleep apnoea: estimation from three randomised controlled trials. *Thorax*. 2019;74(4):390–396.
25. Patel S, Kon SSC, Nolan CM, et al. The Epworth Sleepiness Scale: minimum clinically important difference in obstructive sleep apnea. *Am J Respir Crit Care Med*. 2018;197(7):961–963.
26. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2012;8(5):597–619.
27. Bujang MA. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Archives of Orofacial Sciences*. 2017;12:1–11.
28. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. *Qual Life Res*. 2014;23(6):1713–1720.
29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
30. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346(8982):1085–1087.
31. Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med*. 1999;130(4 Pt 1):289–292.
32. Fong SY, Ho CK, Wing YK. Comparing MSLT and ESS in the measurement of excessive daytime sleepiness in obstructive sleep apnoea syndrome. *J Psychosom Res*. 2005;58(1):55–60.
33. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep*. 1994;17(8):703–710.
34. Aurora RN, Caffo B, Crainiceanu C, Punjabi NM. Correlating subjective and objective sleepiness: revisiting the association using survival analysis. *Sleep*. 2011;34(12):1707–1714.
35. Iftikhar IH, Bittencourt L, Youngstedt SD, et al. Comparative efficacy of CPAP, MADs, exercise-training, and dietary weight loss for sleep apnea: a network meta-analysis. *Sleep Med*. 2017;30:7–14.
36. Bratton DJ, Gaisl T, Schlatzer C, Kohler M. Comparison of the effects of continuous positive airway pressure and mandibular advancement devices on sleepiness in patients with obstructive sleep apnoea: a network meta-analysis. *Lancet Respir Med*. 2015;3(11):869–878.
37. Priest B, Brichard C, Aubert G, Liistro G, Rodenstein DO. Microsleep during a simplified maintenance of wakefulness test. A validation study of the OSLER test. *Am J Respir Crit Care Med*. 2001;163(7):1619–1625.

ACKNOWLEDGMENTS

The authors thank all participants who participated in this study.
Author contributions: Conception and design: FAG, UH, NS, MR, MB, TG, ST, NAS, MK. Funding: MK. Trial conduct: FAG, MR, MB, TG, ST, NAS. Analysis and interpretation of data: FAG, UH, NAS. Drafting the article: FAG. Revising the article for important intellectual content and final approval: All authors.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication October 29, 2019
Submitted in final revised form January 17, 2020
Accepted for publication January 17, 2020

Address correspondence to: Malcolm Kohler, MD, Department of Pulmonology, University Hospital Zurich, Raemistrasse 100, 8091 Zurich, Switzerland; Tel: 41 44 255 38 28; Email: malcolm.kohler@usz.ch

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. This study was funded by grant 2019-01 from Lunge Zurich. The funding sources had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. MK reports grants from University of Zurich and Lunge Zurich during the conduct of the study and grants from Bayer AG, outside the submitted work. TG reports personal fees from Bayer AG, outside the submitted work. All other authors report no conflicts of interest.