Nature and Science of Sleep

Elevate Series

Open Access Full Text Article

ORIGINAL RESEARCH

Test-Retest Reliability of Epworth Sleepiness Scale Score in Patients with Untreated Obstructive Sleep Apnea

Xujun Feng[1,2,*], Yuan Shi[1,*], Ye Zhang[1], Fei Lei[1], Rong Ren[1], Xiangdong Tang[1]

[1]Sleep Medicine Center, Department of Respiratory and Critical Care Medicine, Mental Health Center, West China Hospital, Sichuan University, Chengdu, People's Republic of China; [2]Department of Respiratory and Critical Care Medicine, the First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Rong Ren; Xiangdong Tang, Department of Respiratory and Critical Care Medicine, Sleep Medicine Center, Mental Health Center, West China Hospital, Sichuan University, Chengdu, 610041, People's Republic of China, Tel +86-28-85422733, Email 498880651@qq.com; 2372564613@qq.com

Study Objectives: This study aimed to evaluate the test-retest reliability of the Epworth Sleepiness Scale (ESS) in patients with untreated obstructive sleep apnea (OSA) and investigate the effects of different follow-up techniques and various factors on ESS score discrepancies.

Methods: This study prospectively enrolled participants diagnosed with OSA at West China Hospital of Sichuan University from October 2022 to May 2023. Each participant completed a polysomnography (PSG) and the Chinese version of the ESS. Initial ESS evaluations were performed before the PSG and were reassessed either face-to-face or on telephone within a week. Analysis involved Bland-Altman plots, the intraclass correlation coefficient (ICC), and calculation of mean differences.

Results: We included 382 patients with untreated OSA, averaging 43.52 years old, with a mean body mass index (BMI) of 26.54 kg/m^2 and an average apnea-hypopnea index (AHI) of 47.93 events/hour. The ICC was recorded at 0.820. The signed difference in ESS scores from baseline to follow-up was 1.68 ± 2.93 overall. In OSA patients with a BMI > 28, the difference was 2.39 ± 3.46, while in those with an AHI ≥ 30, it was 1.77 ± 3.27.

Conclusion: This study underscores the significance of repeated ESS testing to improve the reliability of sleepiness evaluations in patients with OSA. Further studies should aim to confirm these findings in a broader demographic and develop refined methods for more precise sleepiness assessments among different OSA groups.

Keywords: Epworth Sleepiness Scale, obstructive sleep apnea, test-retest reliability

Introduction

Obstructive sleep apnea (OSA) is a common, complex, and heterogeneous disorder caused by repeated complete or partial collapse of the upper airway during sleep, resulting in intermittent hypoxia, sleep fragmentation, and autonomic fluctuations during sleep.[1] Globally, approximately 900 million adults aged 30–69 years are affected by mild, moderate, or severe OSA, with China alone reporting a population of 176 million.[2] Excessive daytime sleepiness (EDS), a common symptom of OSA, negatively affects public safety, social interactions, work, mood, cognitive functioning, and quality of life. Recent findings show that OSA with EDS is a more severe subtype of the disorder, resulting in more adverse events and a worse prognosis.[3–5]

Epworth Sleepiness Scale (ESS), a subjective questionnaire, is the most commonly used tool for assessing EDS in clinical practice. Patients are required to answer eight questions, each of which can be given any of four grades (0 for never dozing, 1 for possible mild dozing, 2 for probable dozing, and 3 for very likely dozing or even falling asleep), with total scores ranging from 0 to 24. Generally, ESS scores ≥ 11 are considered to reflect subjective EDS.[6] It has been widely used in OSA screening, assessment, diagnosis, and treatment.[7–9] The degree of change in ESS score is also frequently used as an important endpoint in determining treatment effects in clinical studies.[7,10,11] For example, a 2- or

Received: 10 August 2024
Accepted: 24 December 2024
Published: 31 December 2024

3-point decrease in ESS from baseline after continuous positive airway pressure (CPAP) treatment for OSA is considered the minimal clinically important difference (MCID).[10,12]

The ESS is appreciated for its straightforward nature and ease of use in gauging EDS. Nevertheless, concerns have been raised regarding the dependability and consistency of ESS results across multiple evaluations, notably within clinical cohorts where reliability may deviate from findings in the initial validation.[13–15] Several studies faced constraints due to restricted sample sizes,[13,16,17] while others veered off course from examining ESS stability as they were not designed for that purpose.[18,19] Furthermore, certain investigations faced limitations stemming from statistical methodologies.[16,17,20,21] Although the ESS is designed to reflect long-term average sleep propensity rather than immediate sleepiness, it is recommended that the interval between two measurements should not exceed four weeks to avoid capturing real changes in average sleep propensity.[22] However, only two small-sample studies set the interval between repeated measurements within four weeks.[13,14] Therefore, evaluating the ESS over short intervals could provide a clearer assessment of its reliability and consistency. Moreover, there is a glaring gap in research concerning the test-retest reliability of ESS in individuals with untreated OSA.[23] Given concerns about inconsistent ESS scores in these patients, it is important to consider whether repeated measurements, followed by averaging, similar to blood pressure monitoring, might be implemented. Additionally, it would be useful to investigate whether convenient methods, such as telephone follow-ups, can achieve comparable effectiveness to in-person assessments. Notably, no viable resolutions have been put forth to address this issue.

The main objective of this study was to assess how reliably the ESS can measure daytime sleepiness over short intervals (less than one week) among patients with untreated OSA. Additionally, this study aimed to investigate the factors that affect this consistency. To address potential inconsistencies in ESS measurements, we examined the stability of ESS using different follow-up approaches (eg, "face-to-face" or "telephone"), providing researchers with an easy-to-use and effective means to enhance the reliability of the ESS assessments.

Methods

Participants

This sub-investigation was affiliated with an ongoing prospective study of the pathogenesis and intervention techniques for sleep-breathing disorders approved by the Institutional Review Board of West China Hospital, Sichuan University (No.2022797) and performed in accordance with the Declaration of Helsinki. All participants provided informed consent. This study was performed at the Sleep Center of West China Hospital of Sichuan University.

The research team consecutively screened and enrolled patients who visited the Sleep Center of West China Hospital of Sichuan University between October 2022 and May 2023 and underwent all-night polysomnography (PSG). The inclusion criteria were: (1) Chinese individuals aged 18–75 years. (2) Diagnosis of OSA (ICD-10-CM code: G47.33) fulfilling either of the following two points:[24] a. Apnea-hypopnea index (AHI) ≥ 5/h and accompanied by symptoms of sleepiness, snoring, and apnea, etc. b. AHI ≥ 15/h. The exclusion criteria were: (1) central sleep apnea (CSA; ICD-10-CM code: R06.3; CSA was defined as > 5 episodes of central apnea + central hypoventilation per hour, or the ratio of central apneas + hypoventilation to all apneas + hypoventilation > 50%, or Cheyne–Stokes breathing); (2) disorders that cause chronic sleep disruption (eg, insomnia, pain); (3) current concomitant mental illness (eg, depression) that could affect the accuracy of the questionnaire; (4) comorbidity with other sleep disorders such as narcolepsy, restless legs syndrome, obesity hypoventilation syndrome; (5) combination of uncontrolled acute heart failure, coronary heart disease, chronic obstructive pulmonary disease, and other serious underlying diseases; (6) previous diagnosis of OSA or history of CPAP, or any intervention (such as mandibular orthodontic appliances, hypoglossal nerve stimulation, or sleep position change) during the interval between the two follow-up periods; and (7) inability to independently complete the ESS questionnaire.

Measurements

All patients completed a one-night PSG using the Philips Respironics Alice 6 polysomnographic recorder and Sleepware G3 data analysis software, and all sleep parameters recorded on the PSG were analyzed by senior technicians who were blinded

to the design of the study, according to the international standards of American Academy of Sleep Medicine.[24] The following information about the patients were also collected before sleep monitoring: sex, age, body mass index (BMI), smoking (yes/no), alcohol consumption (yes/no), underlying diseases such as hypertension (yes/no) and diabetes mellitus (yes/no), and a first assessment of ESS. We used the validated Chinese version of the ESS.[25] Patients who met the inclusion criteria were categorized into face-to-face or telephone follow-up groups based on random sampling. The initial ESS assessment occurred in the evening at the Sleep Medicine Center prior to PSG. Subsequently, the ESS score was reassessed via face-to-face or telephone follow-up by experienced sleep medicine physicians (X Feng, Y Shi) within one week (1–6 days) after patients underwent their initial comprehensive sleep evaluation. None of the participants were informed of the purpose of the study.

Sample Size Calculation

A total of 355 participants were required to complete the study protocol, for a power of 90%, $\alpha = 0.05$, and an estimated mean difference in change scores of 0.5 on the ESS, with a projected standard deviation (SD) of 2.9 points. This estimate is based on conservative projections derived from our previous pre-experiment.

Statistical Analysis

Continuous variables are expressed as mean ± SD and categorical variables as numbers (percentages). Comparisons of variables between groups were made using the independent samples *t*-test for continuous variables that conformed to or approximated a normal distribution, the Mann–Whitney *U*-test for those that did not conform to a normal distribution, and the χ^2 test for comparisons between categorical variables.

Intraclass correlation coefficients (ICCs) were used to assess the correlation between the first and second ESS scores.[26] The consistency between the two ESS scores was assessed using Bland-Altman analysis, where the mean difference and the 95% limit of consistency (mean difference ± 1.96 × SD) were calculated as estimates of reproducibility[27] and visualized.[28] Univariate linear regression was used to assess the correlation among the baseline ESS score, sex, age, BMI, smoking, alcohol consumption, hypertension, diabetes mellitus, AHI, nocturnal minimum oxygen saturation, and ESS score difference. In addition, the proportion of participants with ESS score changes ≥ 2, 3, 5, and 7 points between the test and retest was calculated.

A paired *t*-test was performed to assess intragroup differences in ESS comparing follow-up scores to baseline scores. Intergroup differences in ESS changes across follow-ups were analyzed using linear mixed-effects models. The ESS difference served as the dependent variable, with fixed effects for the follow-up group, visit (baseline or within one week), and their interaction (group × visit), which was interpreted as the follow-up effect.

We used SPSS 26.0 for data analysis and GraphPad Prism 9.0 for plotting. $P < 0.05$ was considered to indicate statistical significance.

Results

Participants

Overall, 587 people visited the sleep center for PSG between October 2022 and May 2023, and 401 met the inclusion criteria (Figure 1). Eventually, 382 (mean age, 43.52 ± 12.41, 80.6% men) completed the study, and the mean test-retest interval was 3.60 ± 1.56 days. The number of patients who were followed up face-to-face was 191 (Group A), and those who were followed up by telephone was 191 (Group B). The baseline characteristics and ESS scores of the two groups are shown in Table 1.

Groups A and B did not differ significantly in age, sex, BMI, AHI, lowest SaO_2, or test-retest intervals. Additionally, there were no differences in smoking, alcohol consumption, or the number of patients with comorbid hypertension or diabetes between the two groups. There were no significant differences in baseline, follow-up, mean, or the difference in ESS scores.

Test-Retest Reliability in ESS Scores

The difference in the ESS scores before and after the test was slightly higher in the face-to-face follow-up group (1.70 ± 2.70) compared to the telephone follow-up group (1.66 ± 3.17), but this difference was not significant ($P = 0.967$) (Table S1). Across

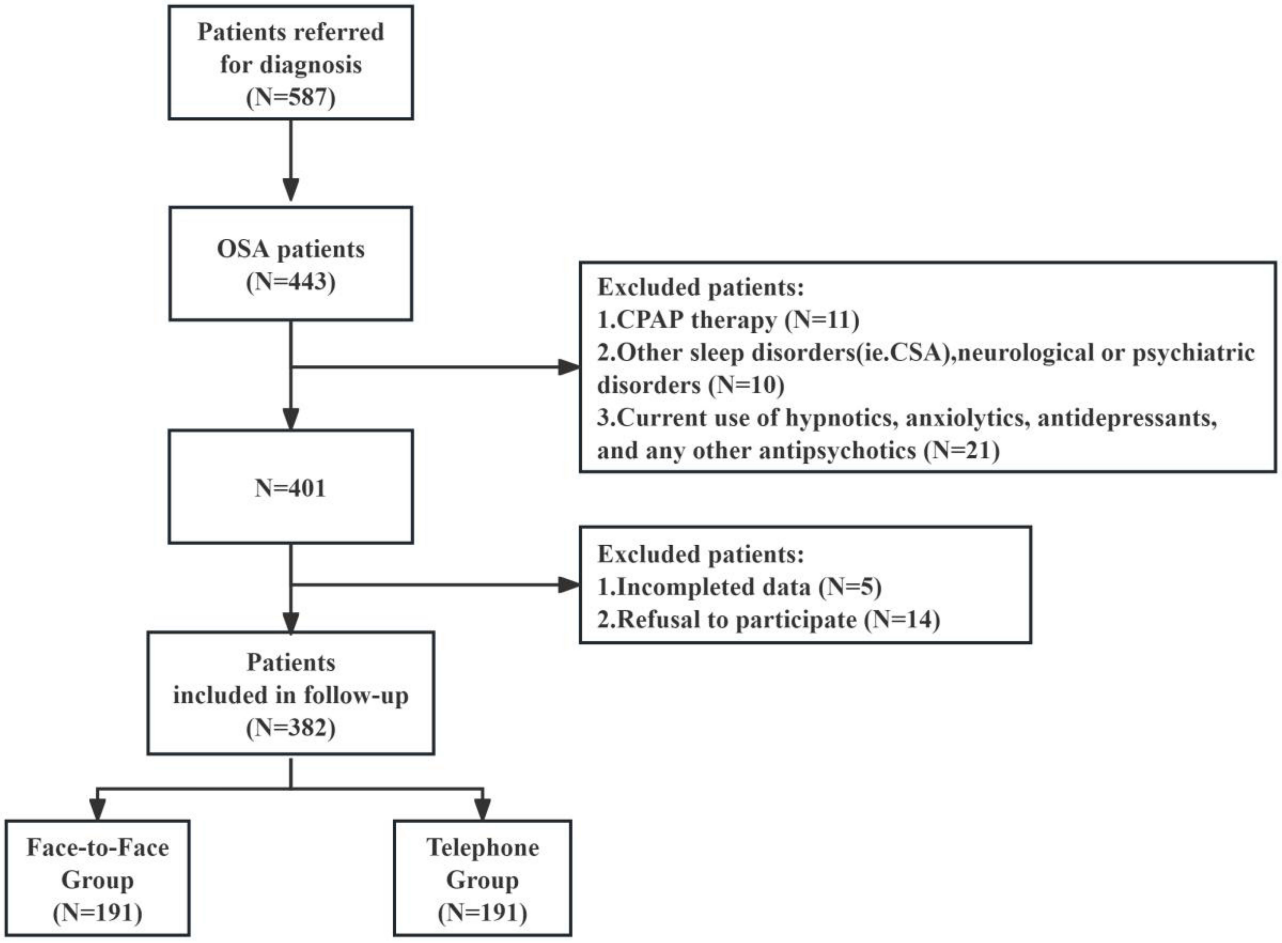

Figure 1 Patient flowchart.

all patients, the percentages with ESS score absolute differences of ≥ 2, 3, 5, and 7 were 66.5%, 49.2%, 17.5%, and 3.9%, respectively. The face-to-face group had a slightly lower percentage of individuals with an ESS score absolute difference of ≥ 3 points (42.3% vs 55.5%, $P = 0.014$).

Figure 2 illustrates the mean ICC values with 95% confidence intervals (CI), both overall and between the groups. The overall ICC for all patients was 0.820 (95% CI: 0.653–0.894). There was no significant difference between the face-to-face and

Table 1 Demographic, Clinical, and Sleep Characteristics of the Study participants

Characteristics	All (n = 382)	Group A (n = 191)	Group B (n = 191)	*P* value
Age, years	43.52 ± 12.41	44.22 ± 12.76	42.83 ± 12.04	0.273
BMI, kg/m²	26.54 ± 4.01	26.40 ± 3.71	26.67 ± 4.29	0.503
Male sex, n (%)	308 (80.6)	150 (78.5)	158 (82.7)	0.300
Hypertension, n (%)	104 (27.2)	46 (24.1)	58 (30.4)	0.168
Diabetes mellitus, n (%)	26 (6.8)	11 (5.8)	15 (7.9)	0.416
Smoking, n (%)	155 (40.6)	79 (41.4)	76 (39.8)	0.755
Alcohol drinking, n (%)	193 (50.5)	91 (47.6)	102 (53.4)	0.260
AHI, events/h	47.93 ± 25.86	47.89 ± 25.10	47.96 ± 26.67	0.979

(*Continued*)

Table 1 (Continued).

Characteristics	All (n = 382)	Group A (n = 191)	Group B (n = 191)	*P* value
$LSaO_2$, %	74.90 ± 12.60	74.09 ± 13.83	75.70 ± 11.23	0.212
Intervals of test-retest	3.60 ± 1.56	3.50 ± 1.55	3.70 ± 1.56	0.212
ESS-first test	7.92 ± 5.51	7.98 ± 5.33	7.86 ± 5.70	0.838
ESS-second test	9.60 ± 5.50	9.68 ± 5.29	9.53 ± 5.72	0.788
ESS-Difference (second - first)	1.68 ± 2.93	1.70 ± 2.70	1.66 ± 3.17	0.967#
ESS-Mean	8.76 ± 5.31	8.83 ± 5.13	8.70 ± 5.49	0.777
ESS absolute difference≥2, n (%)	254 (66.5)	123 (64.3)	131 (68.6)	0.386
ESS absolute difference≥3, n (%)	188 (49.2)	82 (42.3)	106 (55.5)	0.014*
ESS absolute difference≥5, n (%)	67 (17.5)	30 (15.7)	37 (19.4)	0.346
ESS absolute difference≥7, n (%)	15 (3.9)	5 (2.6)	10 (5.2)	0.188

Notes: Group A = face-to-face follow-up; Group B = telephone follow-up; *P<0.05; #: P value for linear mixed-effects model.
Abbreviations: AHI, apnea hypopnea index; BMI, body mass index; $LSaO_2$, %, the lowest oxygen saturation; ESS, Epworth Sleepiness Scale.

telephone groups. Subgroup analyses were conducted to compare ICC values. Patients with severe OSA (AHI≥30) had a lower ICC (ICC=0.784) than those without it (ICC=0.889). Older patients (aged ≥60 years) had a higher ICC than younger patients, though the lower limit of the CI (0.593) was reduced. The higher ICC for older than younger patients was minimal. Patients with obesity (BMI >28) had a relative lower ICC (ICC=0.765), while the female patients demonstrated excellent stability with an ICC of 0.900 (95% CI: 0.832–0.939). However, intragroup comparisons within each subgroup revealed no statistically significant differences, as the 95% CIs overlapped.

A Bland-Altman analysis was performed on all untreated OSA patients (Figure 3a). The y-axis represents the difference between second ESS minus first ESS. The variation in ESS was 1.68 (−3.35 to 8.25). The face-to-face group (1.70 [−3.58 to 6.99]) showed similar results to the telephone group (1.66 [−4.51 to 7.84]) (Figure 3b and c).

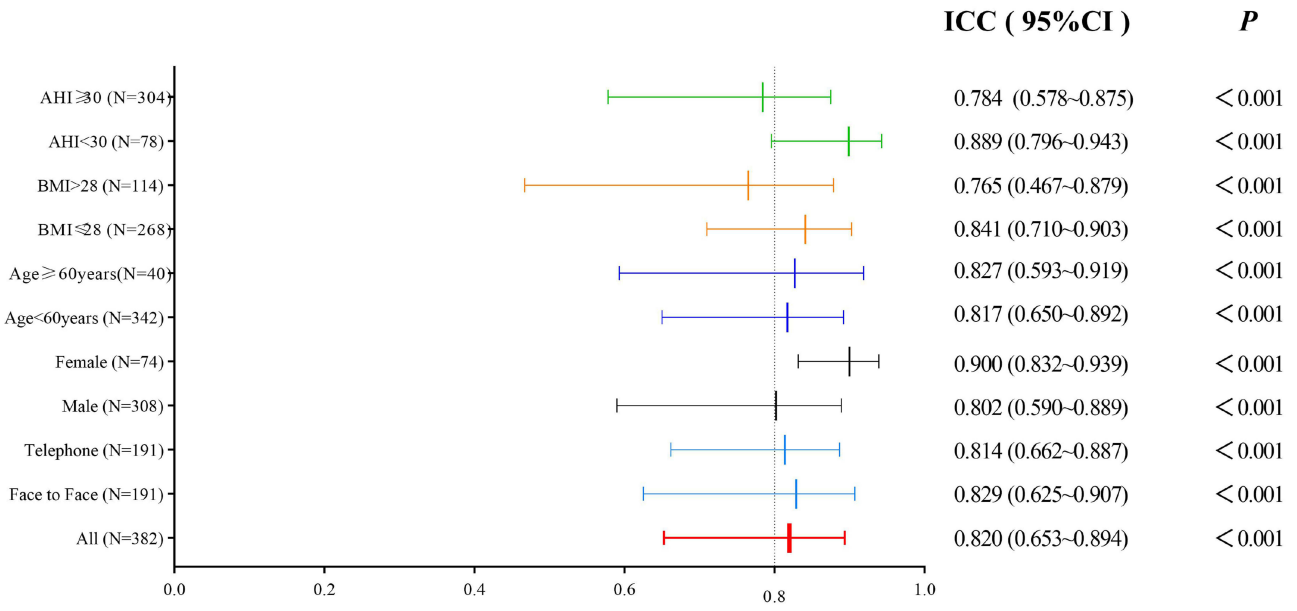

Figure 2 ICCs with the corresponding 95% confidence intervals of the whole cohort and each subgroup.

https://doi.org/10.2147/NSS.S490960

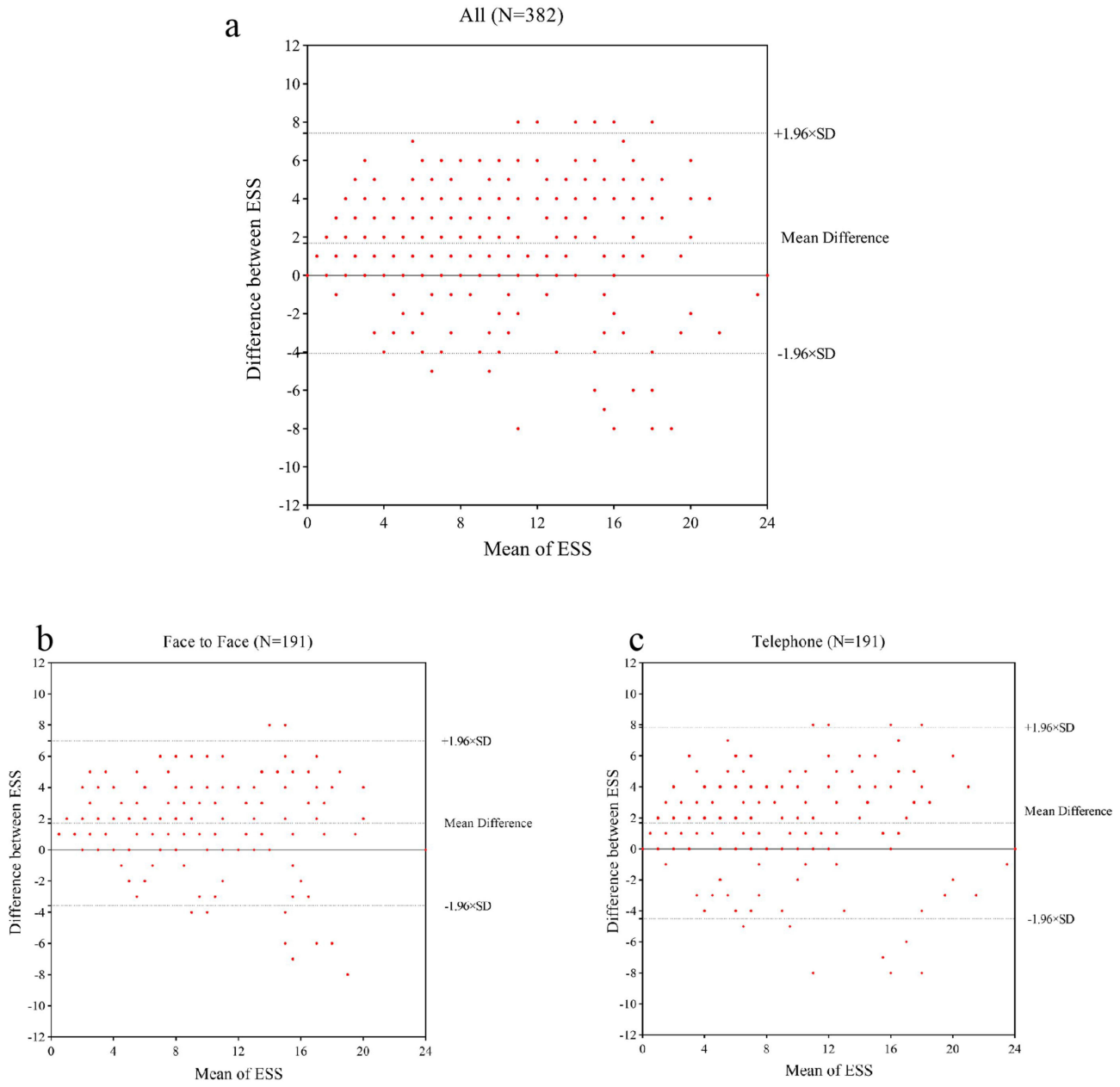

Figure 3 Bland-Altman plots for (**a**) whole cohort, (**b**) face-to-face group, and (**c**) the telephone group. The difference (second minus first) between 2 consecutive ESS scores is plotted against their mean for analyzing levels of accordance between 2 measurements. The Orange dots represent individual test-retest mean ESS scores, while the dashed lines indicate the mean difference and the 95% limit of consistency.

Subgroup analysis was performed based on OSA severity (Figure S1), BMI (Figure S2), age (Figure S3), and sex (Figure S4). Patients with severe OSA and obesity showed a variation of 1.77 (−4.13 to 7.67) and 2.39 (−4.39 to 9.18), respectively. The variation observed in female OSA patients was 0.82 (−3.69 to 5.34).

Influence of Baseline Variables on the Variation of ESS

Table 2 illustrates the relationship between various baseline characteristics and the signed difference in sequential ESS scores (second minus first), as analyzed through univariate linear regression. Factors that influenced the stability of repeated ESS measurements included BMI, sex, AHI, and baseline ESS.

Table 2 Univariate Linear Regression of Influence of Baseline Parameters on ESS Differences

Parameter	Coefficient	95% CI	R squared	*P* value
Age	0.001	−0.022 to 0.025	0	0.909
BMI	0.10	0.027 to 0.173	0.019	0.008**
Male sex	1.065	0.327 to 1.804	0.021	0.005**
Hypertension	0.475	−0.186 to 1.136	0	0.159
Diabetes mellitus	1.000	−0.167 to 2.167	0.007	0.093
Smoking	0.316	−0.284 to 0.916	0.003	0.301
Alcohol drinking	−0.061	−0.651 to 0.528	0	0.838
AHI	0.014	0.003 to 0.026	0.016	0.014*
$LSaO_2$, %	−0.022	−0.045 to 0.002	0.009	0.068
Intervals of follow-up	−0.076	−0.265 to 0.114	0.002	0.433
ESS at baseline	−0.142	−0.194 to −0.091	0.072	<0.001**

Notes: *P<0.05; **P<0.01.
Abbreviations: AHI, apnea hypopnea index; BMI, body mass index; $LSaO_2$, %, the lowest oxygen saturation; ESS, Epworth Sleepiness Scale; CI, confidence interval.

Discussion

Our study assessed the short-term reliability of the ESS in individuals with untreated OSA. We meticulously evaluated the test-retest reliability of the ESS in 382 patients diagnosed with OSA at the Sleep Medicine Center of West China Hospital. Statistically, the ESS, a common measure of EDS, demonstrated an ICC of 0.820, which is greater than the cut-off value of 0.8, indicating substantial reliability.[29] The reliability was consistent for both face-to-face (ICC = 0.829) and telephone follow-ups (ICC = 0.814), suggesting that the administration method does not significantly impact this stability. Given its consistent performance across follow-up methods, overall, the ESS proves to be a reliable tool for assessing daytime sleepiness in untreated OSA patients. This reliability increases its value in routine clinical practice, helping monitor the severity of sleepiness and guiding treatment decisions in OSA management. However, from the perspective of clinical practice, the mean difference between repeated ESS measurements was 1.68 ± 2.93. Although relatively small, this difference is significant when considering the MCID for the ESS.[10,12] The study also found that patients with severe OSA and those with obesity had moderate stability in ESS scores, with ICC of 0.784 and 0.765, respectively. This reduced reliability in specific subgroups is crucial, as it underscores potential variability in EDS assessments among these patients. Further analysis revealed that BMI, male sex, and the AHI were positively correlated with variations in ESS scores. Conversely, baseline ESS scores showed a negative correlation with these variations. These findings indicate that specific demographic and clinical factors can affect the reliability of ESS scores, requiring more nuanced interpretation in clinical practice.

Johns,[6,15] the developer of the ESS, identified strong reliability coefficients in healthy populations, confirming the scale's robustness. The reliability of the ESS found in this study is consistent with previous research. For example, studies by Grewe et al[13] and Veqar et al[30] reported ICCs ranging from 0.73 to 0.86, indicating moderate reliability of the ESS in clinical settings. The short interval might contribute to higher ICCs observed, as longer periods typically introduce more variability in measurements. However, our study introduces a new perspective by highlighting the variability among patients with untreated OSA. While this study confirmed the statistically good stability of the ESS, the mean difference between the two measurements was approximately 1.7 points. Notably, among patients with obesity, this mean difference increased to 2.39 points. Such a high value is not a good sign for the reliability despite the other methods, particularly in the context of OSA, where a large proportion of patients are obese. To our knowledge, a change

of 2 points on the ESS is clinically significant, as numerous studies on OSA treatment consider a 2–3-point change to be an indicator of significant effectiveness.[10,12] Additionally, a recent meta-analysis revealed that the placebo effect of sham-CPAP on the ESS was equivalent to a change of −1.55 points.[31]

Previous studies, constrained by limited sample sizes and study designs,[13,14,32] have not adequately addressed this issue from both a statistical and clinical practice perspective. This gap underscores the significance of our findings. In our subgroup analysis, we discovered that, particularly for patients with severe OSA and obesity, a single ESS measurement may be unreliable. While the exact mechanism remains unclear, we speculated that this reduced reliability might be due to heightened symptom variability and psychological factors affecting the perception of sleepiness, and to the subjective nature of the ESS. For instance, individuals with obesity might encounter more significant fluctuations in daily activities and mood,[33] which can impact their ESS scores. Additionally, more severe OSA exacerbates sleep fragmentation, leading to fluctuations in sleep quality and daytime performance. Simultaneously, the psychological impact of severe OSA, such as anxiety and depression, influences how patients perceive and communicate their symptoms.

Our findings may have meaningful clinical implications. While the ESS is generally reliable, the identified variability in patients with severe OSA and obesity has crucial implications for clinical practice. Clinicians should exercise caution when interpreting ESS scores in these groups, as the scores may fluctuate more than in other patients. This variability also highlights the importance of considering individual patient factors when using the ESS to assess sleepiness and developing tailored assessment strategies for these specific populations.

Averaging the ESS scores twice in a week may provide more reliable results in this population. This approach is akin to the method used in clinical practice for measuring blood pressure, where the average of two or three readings is typically taken.[34,35] Our study's randomization into face-to-face and telephone follow-ups offers unique insights into how the mode of administration affects reliability, an aspect not extensively covered in previous research. The similar ICC and mean difference of ESS scores for both follow-up methods indicate that ESS can be reliably administered through different modes, enhancing its usefulness in various clinical and research settings. Developing modified versions of the ESS for patients with OSA could further improve reliability. This could include additional questions addressing mood or a broader range of daily activities, providing a more comprehensive understanding of these patients' experiences. Training clinicians on the proper administration and interpretation of the ESS, with a focus on individual patient factors, could further enhance the scale's utility.

This study has several limitations. First, the study involved 382 patients from a single sleep center, which may not be representative of all patients with OSA, especially those from different geographic, cultural, or socioeconomic backgrounds. This might limit the generalizability of the findings. Future studies with larger, more diverse populations are needed to confirm these findings. Second, this study repeated ESS measurements within a one-week interval. However, until now, there is no consensus on the optimal follow-up interval. Third, measurement errors could arise from patients' recall bias or misunderstanding of the ESS questions. Despite efforts to standardize the administration of the ESS, individual differences in comprehension and recall could influence the reported scores, thereby affecting the observed reliability. Fourth, although ICC may be more appropriate than Pearson's correlation test in assessing the stability of repeated measures of subjective scales,[36] we must acknowledge the limitations of this statistical method itself.[26] ICC assumes that the variance components (between-subject and within-subject) are homogeneous across all subjects. If the variability differs significantly between subjects, the ICC value might be misleading, overestimating or underestimating the true reliability.[37] Moreover, there is no universally accepted threshold for interpreting ICC values. While some studies consider an ICC greater than 0.75 to indicate good reliability, others set the threshold at 0.8 or even 0.9, highlighting the lack of consensus in this regard.[38,39] Fifth, the absence of a control group of healthy individuals without OSA is an important limitation. This omission prevented a comparison of ESS reliability between OSA patients and healthy individuals, which could provide insight into whether ESS reliability is primarily influenced by OSA or other factors, such as individual response biases. Sixth, during one week, the reliability can vary due to a concomitant event that may considerably change the ESS both a baseline and for the second measurement. Furthermore, although a one-week interval was selected to minimize external influences and ensure comparability, this relatively short timeframe may not capture long-term variations in sleepiness. Extended intervals could better reflect fluctuations due to lifestyle changes, adaptation to OSA symptoms, or other dynamic factors, thus providing a more comprehensive evaluation of ESS reliability. Future research should address these limitations by incorporating a healthy control group, evaluating reliability over extended

intervals, and investigating the impact of specific confounders such as comorbidities (eg, depression, anxiety), lifestyle factors, medication use, and variations in OSA severity.

Conclusions

In conclusion, this study revealed that the ESS is a moderately reliable tool to assess sleepiness in patients with untreated OSA. Reassessing the ESS score within one week—either through face-to-face or telephone follow-up—and averaging the results may offer a more precise evaluation of sleepiness, particularly in patients with severe OSA and obesity. Future research should aim to validate these findings across broader populations and develop strategies to enhance the reliability of sleepiness assessments in diverse OSA groups.

Brief Summary

Current Knowledge/Study Rationale: This study aimed to assess the Epworth Sleepiness Scale's reliability for measuring daytime sleepiness in patients with untreated obstructive sleep apnea over short intervals. The factors influencing this reliability were also explored.

Study Impact: The Epworth Sleepiness Scale has a moderate reliability for assessing sleepiness in patients with untreated obstructive sleep apnea. However, the scale's reliability diminishes in patients with obesity and severe obstructive sleep apnea; this may be due to heightened symptom variability and psychological factors affecting the perception of sleepiness in these patients.

Abbreviations

AHI, apnea-hypopnea index; BMI, body mass index; CI, confidence intervals; CPAP, continuous positive airway pressure; CSA, central sleep apnea; EDS, Excessive daytime sleepiness; ESS, Epworth Sleepiness Scale; ICC, intraclass correlation coefficient; MCID, minimal clinically important difference; OSA, obstructive sleep apnea; PSG, polysomnography; SD, standard deviation.

Acknowledgments

The authors wanted to thank all the patients for their participation in this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting or writing, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for the contents of this article.

Funding

This work was supported by the Ministry of Science and Technology of the People's Republic of China (STI2030-Major Projects 2021ZD0201900) and was supported by National Natural Science Foundation of China (82341249, 82170099).

Disclosure

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers*. 2015;1:15015. doi:10.1038/nrdp.2015.15
2. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687–698. doi:10.1016/S2213-2600(19)30198-5
3. Li X, Huang H, Xu H, et al. Excessive daytime sleepiness, metabolic syndrome, and obstructive sleep apnea: two independent large cross-sectional studies and one interventional study. *Respir Res*. 2019;20(1):276. doi:10.1186/s12931-019-1248-y
4. Xie J, Sert Kuniyoshi FH, Covassin N, et al. Excessive daytime sleepiness independently predicts increased cardiovascular risk after myocardial infarction. *J Ame Heart Assoc*. 2018;7(2). doi:10.1161/JAHA.117.007221.

5. Mazzotti DR, Keenan BT, Lim DC, Gottlieb DJ, Kim J, Pack AI. Symptom subtypes of obstructive sleep apnea predict incidence of cardiovascular outcomes. *Am J Respir Crit Care Med*. 2019;200(4):493–506. doi:10.1164/rccm.201808-1509OC
6. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14(6):540–545. doi:10.1093/sleep/14.6.540
7. Furlow B. SAVE trial: no cardiovascular benefits for CPAP in OSA. *Lancet Respir Med*. 2016;4(11):860. doi:10.1016/S2213-2600(16)30300-9
8. Jonas DE, Amick HR, Feltner C, et al. Screening for obstructive sleep apnea in adults: evidence report and systematic review for the us preventive services task force. *JAMA*. 2017;317(4):415–433. doi:10.1001/jama.2016.19635
9. Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. *Lancet*. 2014;383(9918):736–747. doi:10.1016/S0140-6736(13)60734-5
10. 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. doi:10.1136/thoraxjnl-2018-211959
11. Craig S, Pépin JL, Randerath W, et al. Investigation and management of residual sleepiness in CPAP-treated patients with obstructive sleep apnoea: the European view. *European Resp Rev*. 2022;31(164):210230. doi:10.1183/16000617.0230-2021
12. 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. doi:10.1164/rccm.201704-0672LE
13. 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. doi:10.5664/jcsm.8350
14. Lee JL, Chung Y, Waters E, Vedam H. The Epworth sleepiness scale: reliably unreliable in a sleep clinic population. *J Sleep Res*. 2020;29(5):e13019. doi:10.1111/jsr.13019
15. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*. 1992;15(4):376–381. doi:10.1093/sleep/15.4.376
16. 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. doi:10.5664/jcsm.7108
17. Chung KF. Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosomatic Res*. 2000;49(5):367–372. doi:10.1016/S0022-3999(00)00186-0
18. Rosenberg R, Babson K, Menno D, et al. Test-retest reliability of the Epworth Sleepiness Scale in clinical trial settings. *J Sleep Res*. 2022;31(2):e13476. doi:10.1111/jsr.13476
19. Knutson KL, Rathouz PJ, Yan LL, Liu K, Lauderdale DS. Stability of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness questionnaires over 1 year in early middle-aged adults: the CARDIA study. *Sleep*. 2006;29(11):1503–1506. doi:10.1093/sleep/29.11.1503
20. 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.
21. Schober P, Mascha EJ, Vetter TR. Statistics from A (Agreement) to Z (z score): a guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. *Anesthesia Analg*. 2021;133(6):1633–1641. doi:10.1213/ANE.0000000000005773
22. Streiner D, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use (5th edition). *Aust N. Z. J Public Health*. 2016;40(3):294–295. doi:10.1111/1753-6405.12484
23. Scharf MT. Reliability and efficacy of the Epworth Sleepiness Scale: is there still a place for it? *Nat Sci Sleep*. 2022;14:2151–2156. doi:10.2147/NSS.S340950
24. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2017;13(3):479–504. doi:10.5664/jcsm.6506
25. Peng LL, Li JR, Sun JJ, et al. Reliability and validity of the simplified Chinese version of Epworth sleepiness scale. *Chinese J Otorhinolaryngol Head Neck Surg*. 2011;46(1):44–49.
26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428. doi:10.1037/0033-2909.86.2.420
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310. doi:10.1016/S0140-6736(86)90837-8
28. Gerke O. Reporting standards for a Bland-Altman agreement analysis: a review of methodological reviews. *Diagnostics*. 2020;10(5):334.
29. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Meth Med Res*. 1998;7(3):301–317. doi:10.1177/096228029800700306
30. Veqar Z, Hussain ME. Psychometric analysis of Epworth Sleepiness Scale and its correlation with Pittsburgh sleep quality index in poor sleepers among Indian university students. *Int j Adolescent Med Health*. 2018;31(2). doi:10.1515/ijamh-2016-0151
31. Labarca G, Montenegro R, Oscullo G, et al. Placebo response in objective and subjective measures of hypersomnia in randomized clinical trials on obstructive sleep apnea. A systematic review and meta-analysis. *Sleep Med Rev*. 2023;67:101720. doi:10.1016/j.smrv.2022.101720
32. Taylor E, Zeng I, O'Dochartaigh C. The reliability of the Epworth Sleepiness Score in a sleep clinic population. *J Sleep Res*. 2019;28(2):e12687. doi:10.1111/jsr.12687
33. Hamer M, Batty GD, Kivimaki M. Risk of future depression in people who are obese but metabolically healthy: the English longitudinal study of ageing. *Mol Psychiatry*. 2012;17(9):940–945. doi:10.1038/mp.2012.30
34. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: the Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. *J Hypertension*. 2018;36(10):1953–2041. doi:10.1097/HJH.0000000000001940
35. Zhang DY, Cheng YB, Guo QH, et al. Treatment of masked hypertension with a Chinese herbal formula: a randomized, placebo-controlled trial. *Circulation*. 2020;142(19):1821–1830. doi:10.1161/CIRCULATIONAHA.120.046685
36. Yen M, Lo LH. Examining test-retest reliability: an intra-class correlation approach. *Nursing Res*. 2002;51(1):59–62. doi:10.1097/00006199-200201000-00009
37. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Soc Admin Pharm*. 2013;9(3):330–338. doi:10.1016/j.sapharm.2012.04.004
38. Shen WQ, Yao L, Wang XQ, Hu Y, Bian ZX. Quality assessment of cancer cachexia clinical practice guidelines. *Cancer Treat Rev*. 2018;70:9–15. doi:10.1016/j.ctrv.2018.07.008
39. Xue C, Yuan J, Lo GG, et al. Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. *Quantitative Imaging Med Surg*. 2021;11(10):4431–4460. doi:10.21037/qims-21-86