RESEARCH ARTICLE

Test–retest reliability of the Epworth Sleepiness Scale in clinical trial settings

Russell Rosenberg[1,2] [iD] | Kimberly Babson[3] | Diane Menno[3] | Susan Morris[3] | Michelle Baladi[3] | Danielle Hyman[3] | Jed Black[4]

[1]NeuroTrials Research, Atlanta, GA, USA

[2]Atlanta School of Sleep Medicine, Atlanta, GA, USA

[3]Jazz Pharmaceuticals, Palo Alto, CA, USA

[4]Stanford Center for Sleep Sciences and Medicine, Palo Alto, CA, USA

Correspondence
Russell Rosenberg, Neurotrials Research, Inc., 5887 Glenridge Drive, Suite 400, Atlanta, GA 30328, USA.
Email: rrosenberg@neurotrials.com

Summary

The present analysis examined the test–retest reliability of the Epworth Sleepiness Scale in participants with excessive daytime sleepiness associated with narcolepsy or obstructive sleep apnea in three clinical trials. Intraclass correlation coefficient estimates for Epworth Sleepiness Scale scores from two solriamfetol 12-week placebo-controlled trials (one narcolepsy, one obstructive sleep apnea) and one long-term open-label extension trial (narcolepsy or obstructive sleep apnea) were calculated using postbaseline time-point pairs for the overall population in each trial, by treatment, and by primary obstructive sleep apnea therapy adherence. In the 12-week narcolepsy trial, intraclass correlation coefficients (95% confidence intervals) were 0.83 (0.79, 0.87) for weeks 4 and 8 ($n = 199$), 0.87 (0.83, 0.90) for weeks 8 and 12 ($n = 196$), and 0.81 (0.76, 0.85) for weeks 4 and 12 ($n = 196$). In the 12-week obstructive sleep apnea trial, intraclass correlation coefficients (95% confidence intervals) were 0.74 (0.69, 0.78) ($n = 416$), 0.80 (0.76, 0.83) ($n = 405$), and 0.74 (0.69, 0.78) ($n = 405$), respectively. In the open-label extension trial, intraclass correlation coefficients (95% confidence intervals) were 0.82 (0.79, 0.85) for weeks 14 and 26/27 ($n = 495$), 0.85 (0.82, 0.87) for weeks 26/27 and 39/40 ($n = 463$), and 0.78 (0.74, 0.81) for weeks 14 and 39/40 ($n = 463$). Placebo/solriamfetol treatment or adherence to primary obstructive sleep apnea therapy did not affect reliability. In conclusion, across three large clinical trials of participants with narcolepsy or obstructive sleep apnea, Epworth Sleepiness Scale scores demonstrated a robust acceptable level of test–retest reliability in evaluating treatment response over time.

KEYWORDS
JZP-110, repeated measurements, reproducibility, Sunosi, surveys and questionnaire, wake-promoting agents

---

1 | INTRODUCTION

The Epworth Sleepiness Scale (ESS) is a patient-reported questionnaire that measures excessive daytime sleepiness (EDS) by assessing situational sleep propensities (Johns, 1991). Specifically, the ESS is composed of eight items that assess the likelihood of falling asleep in real-world situations, such as reading, watching television, or driving. Each item is scored from zero to 3, for a total score of zero to 24, with higher scores indicating a greater severity of EDS. Scores of ≤10 are commonly considered within the normal range, whereas scores of >10 indicate pathological EDS (Johns & Hocking, 1997; Johns, 1991). Data from patients with obstructive sleep apnea (OSA) suggest that changes of 2–3 points may be considered the minimum clinically important difference on the ESS (Crook et al., 2019; Patel et al., 2018).

The ESS is widely used in sleep research, clinical trials, and clinical practice. In clinical trials, the ESS is often used to evaluate the effects of treatment intervention on EDS in several disease states, including narcolepsy and OSA. For instance, the efficacy and regulatory approvals of several wake-promoting agents, such as solriamfetol, modafinil, armodafinil, and pitolisant, has been supported by reductions (improvements) in ESS scores (Black & Hirshkowitz, 2005; Black et al., 2010; Dauvilliers et al., 2013, 2020; Harsh et al., 2006; Malhotra et al., 2020; Schweitzer et al., 2019; Szakacs et al., 2017; Thorpy et al., 2019; US Modafinil in Narcolepsy Multicenter Study Group, 2000). Despite this common utility, there are few studies that have evaluated the test–retest reliability of the ESS (Kendzerska et al., 2014).

The test–retest reliability of the ESS has primarily been investigated in healthy, community-based samples (Ahmed et al., 2014; Johns, 1992; Knutson et al., 2006). A few studies (Campbell et al., 2018; Lee et al., 2020; Nguyen et al., 2006; Rozgonyi et al., 2021; Taylor et al., 2019; Walker et al., 2020) have evaluated the reliability of the ESS in sleep clinic patients with suspected sleep disorders by retrospective chart review with conflicting results. In real-world clinic settings, multiple variables could change between assessments, such as the setting of the assessment (primary care setting versus sleep specialist setting), treatment interventions, caffeine use, and concomitant medications. Such factors could impact EDS and lead to greater variability in the ESS scores. These settings are not ideal for evaluating the test–retest reliability of the ESS in relation to its use as an outcome measure in a clinical trial setting. Therefore, it is necessary to examine the test–retest reliability of the ESS within a controlled clinical trial setting, in which multiple factors, such as those previously noted, would be uniform. However, there is a paucity of data ascertaining whether the ESS is reliable in a clinical trial setting and, furthermore, whether the reliability is observed in patients with sleep disorders other than suspected OSA.

Solriamfetol, a dopamine and noradrenaline re-uptake inhibitor, is approved in the United States and European Union to improve wakefulness in adult patients with EDS associated with narcolepsy (75–150 mg/day) or OSA (37.5–150 mg/day) (Sunosi™ (solriamfetol) tablets Prescribing Information, 2019; Sunosi™ (solriamfetol) tablets Summary of Product Characteristics, 2020). In two randomised, placebo-controlled, phase III trials and one long-term open-label extension (OLE) trial evaluating the effects of solriamfetol in participants with EDS associated with narcolepsy or OSA, ESS scores were included as a primary or co-primary outcome measure (Malhotra et al., 2020; Schweitzer et al., 2019; Thorpy et al., 2019). The large sample sizes and structured nature of these studies provided an opportunity to assess the test–retest reliability of the ESS in a clinical sample in a clinical trial setting.

The aim of the present analysis was to examine the test–retest reliability of the ESS in participants with narcolepsy or OSA in a clinical trial setting, using the intraclass correlation coefficient (ICC) method (US Department of Health and Human Services, 2009). In addition, this analysis evaluated whether certain factors that might affect EDS, including treatment (with placebo or solriamfetol) and adherence to primary OSA therapy (i.e. adherence or nonadherence), impact the reliability of the ESS.

2 | METHODS

2.1 | Study design

The present analysis includes data from phase III research from the solriamfetol clinical trial programme. This included two 12-week, randomised, double-blind, placebo-controlled, phase III clinical trials: one in participants with narcolepsy (NCT02348593/EudraCT 2014-005487-15) (Thorpy et al., 2019); one in participants with OSA (NCT02348606/EudraCT 2014-005514-31) (Schweitzer et al., 2019), and one OLE trial in participants with narcolepsy or OSA (NCT02348632/EudraCT 2014-005489-31) (Malhotra et al., 2020).

All studies were approved by institutional review boards or ethics committees at each institution and were performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. Complete descriptions of the study methods and primary results have been published previously (Malhotra et al., 2020; Schweitzer et al., 2019; Thorpy et al., 2019) and the methods are briefly summarised below.

2.2 | Participants

For the 12-week trials, eligible participants were adults (aged 18–75 years) diagnosed with narcolepsy (Type 1 or Type 2) or OSA and with ESS scores of ≥10. Additional key inclusion criteria included baseline mean sleep latency <25 min (narcolepsy) or <30 min (OSA) on the Maintenance of Wakefulness Test (MWT), and usual nightly total sleep time of ≥6 hr. Additional inclusion criteria were that participants with OSA were required to have current or history of prior (or attempted) use of a primary OSA therapy (i.e. to treat the underlying airway obstruction), including positive airway pressure, mandibular advance device, or surgical intervention. Participants without current primary OSA therapy use or a history of a successful

surgical intervention to treat the underlying obstruction were required to have tried to use a primary OSA therapy for ≥1 month, with at least one documented adjustment to the therapy. At study entry, participants were instructed to maintain the same level of use of primary OSA therapy throughout the study. Key exclusion criteria included usual bedtime later than 1:00 a.m., night-time or variable shift work, or any other clinically relevant medical, behavioural, or psychiatric disorder associated with EDS. Concomitant treatment with other medications that may affect the evaluation of EDS was not permitted.

For the OLE trial, participants with narcolepsy or OSA who had previously completed a phase II or phase III clinical trial of solriamfetol were eligible (Bogan et al., 2015; Ruoff et al., 2016; Schweitzer et al., 2019; Strollo et al., 2019; Thorpy et al., 2019). Due to differences in time between prior study completion and enrolment in the OLE trial, there were two groups. Group A enrolled in the OLE trial immediately after completion of one of the 12-week phase III trials. Group B historically completed one of several other solriamfetol studies and was subsequently enrolled in the OLE trial.

2.3 | Treatment

In the 12-week trials, participants were randomised to receive placebo or solriamfetol 37.5 (OSA only), 75, 150, or 300 mg once daily for 12 weeks. In the OLE trial, solriamfetol treatment was initiated at 75 mg and titrated to 75, 150, or 300 mg during a 2-week titration phase. The titration phase was followed by an open-label maintenance phase (75, 150, or 300 mg), with a total study duration of 40 weeks (Group A) or 52 weeks (Group B).

2.4 | ESS assessments

In all three trials, the ESS was administered at the investigative sites (sleep clinics) for all assessments. In the 12-week trials, the ESS was assessed at baseline and at weeks 1, 4, 8, and 12. In the OLE trial, the ESS was assessed at baseline and at weeks 2, 14, 27, and 40 (Group A) or at weeks 2, 14, 26, 39, and 52 (Group B). Participants were instructed to complete the ESS based on the level of sleepiness they experienced over the past week (7-day recall period) (Broderick et al., 2013; Plazzi et al., 2018).

2.5 | Statistical analysis

For the 12-week trials, data were analysed for the modified intent-to-treat (mITT) population, which was used for the primary efficacy analyses in these studies (Schweitzer et al., 2019; Thorpy et al., 2019) and was defined as participants who received one or more doses of study medication and had baseline and one or more post-baseline assessments. For the OLE trial, data were analysed for the

safety population, defined as participants who received one or more doses of solriamfetol. The ICC estimates for ESS scores were calculated using postbaseline time-point pairs. In the 12-week trials, the time-point pairs were weeks 4 and 8, weeks 8 and 12, and weeks 4 and 12. In the OLE trial, the time-point pairs were weeks 14 and 26/27, weeks 26/27 and 39/40, and weeks 14 and 39/40 (week 52 was not included, as only Group B had data at this time-point). All analyses included participants who had data at both visits for each time-point pair.

The ICC estimates and 95% confidence intervals (CIs) were calculated for the overall population in each of the three trials and for the populations of the two 12-week trials combined. For each 12-week trial, the ICC estimates were also calculated by treatment (placebo or combined solriamfetol [all doses]). For participants with OSA (the full population of the 12-week OSA trial and the OSA subgroup of the OLE trial), ICC estimates were calculated by adherence or non-adherence to primary OSA therapy. Participants were categorised as adherent to primary OSA therapy if they had use of positive airway pressure therapy for ≥4 hr/night on ≥70% of nights, use of an oral appliance on ≥70% of nights, or receipt of an effective surgical intervention. Participants were categorised as non-adherent if they had device use at a frequency/duration less than that described above, no use of a device at all, or a surgical intervention deemed no longer effective.

The ICC estimates and 95% CIs were calculated for each subsample using a two-way mixed-effects model, according to the method of Shrout and Fleiss (Shrout & Fleiss, 1979).

3 | RESULTS

3.1 | Participant population

In the 12-week trials, the mITT populations included 231 participants with narcolepsy (placebo, $n = 58$; combined solriamfetol, $n = 173$) and 459 participants with OSA (placebo, $n = 114$; combined solriamfetol, $n = 345$). In the OLE trial, the safety population included 643 participants.

3.2 | Participant demographics

Across all three trials, the majority of participants were White, not Hispanic or Latino, and primarily enrolled at sites in North America. In the 12-week trial in participants with narcolepsy (mITT population), the majority of participants were female, mean age was ~36 years, and mean body mass index (BMI) was ~28 kg/m^2 (Table 1). In the 12-week trial of OSA (mITT population), the majority of participants were male, the mean age was ~54 years, and the mean BMI was ~33 kg/m^2 (Table 1). In the OLE trial (safety population), 52% of participants were male, the mean age was ~49 years, and the mean BMI was ~32 kg/m^2 (baseline data for the OLE trial have been previously reported (Malhotra et al., 2020)).

TABLE 1 Demographic and baseline clinical characteristics[a]

Characteristic	12-week study – OSA		12-week study – Narcolepsy		OLE
	Placebo (n = 114)	Combined solriamfetol (all doses) (n = 345)	Placebo (n = 58)	Combined solriamfetol (all doses) (n = 173)	Combined solriamfetol (all doses) (n = 643)[b]
Age, years, mean (SD)	54.0 (11.5)	53.8 (10.8)	36.2 (15.2)	36.2 (12.4)	49.3 (14.2)
Gender, n (%)					
Male	73 (64.0)	214 (62.0)	24 (41.4)	57 (32.9)	337 (52.4)
Female	41 (36.0)	131 (38.0)	34 (58.6)	116 (67.1)	306 (47.6)
BMI, kg/m², mean (SD)	33.1 (5.3)	33.3 (5.3)	29.3 (5.8)	28.0 (5.8)	31.7 (5.9)
Race, n (%)					
American Indian or Alaska Native	1 (0.9)	0	0	2 (1.2)	n/a
Asian	4 (3.5)	13 (3.8)	0	6 (3.5)	15 (2.3)
Black or African American	26 (22.8)	61 (17.7)	10 (17.2)	23 (13.3)	109 (17.0)
Native Hawaiian or Other Pacific Islander	1 (0.9)	1 (0.3)	0	1 (0.6)	n/a
White	82 (71.9)	266 (77.1)	46 (79.3)	138 (79.8)	506 (78.7)
Multiple or Other	0	4 (1.2)	2 (3.4)	3 (1.7)	13 (2.0)
MWT sleep latency, min, mean (SD)[c]	12.6 (7.1)	12.5 (7.4)	6.2 (5.7)	8.0 (5.8)	n/a
ESS score, mean (SD)	15.6 (3.3)	15.1 (3.3)	17.3 (2.9)	17.2 (3.3)	15.9 (Group A) 15.9 (Group B)

BMI, body mass index; ESS, Epworth Sleepiness Scale; mITT, modified intent to treat; MWT, Maintenance of Wakefulness Test; n/a, not applicable; OLE, open-label extension; OSA, obstructive sleep apnea; SD, standard deviation.

[a]mITT population.

[b]Data from Malhotra 2019.

[c]For baseline mean sleep latency on MWT, OSA placebo, n = 111; OSA solriamfetol, n = 339; narcolepsy placebo, n = 57; narcolepsy solriamfetol, n = 170.

3.3 | Test–retest reliability of ESS scores in the 12-week and 1-year trials (pooled data)

In the overall study populations, ICC estimates (95% CIs) ranged from 0.78 (0.75, 0.81) for weeks 4 and 12 to 0.84 (0.82, 0.87) for weeks 8 and 12 in the 12-week trials (pooled data); and from 0.78 (0.74, 0.81) for weeks 14 and 39/40 to 0.85 (0.82, 0.87) for weeks 26/27 and 39/40 in the OLE trial (Figure 1).

3.4 | Test–retest reliability of ESS scores in the 12-week trials (by indication)

In the individual 12-week trials, the ICC estimates (95% CI) ranged from 0.81 (0.76, 0.85) for weeks 4 and 12 to 0.87 (0.83, 0.90) for weeks 8 and 12 for participants with narcolepsy and from 0.74 (0.69, 0.78) for weeks 4 and 8 to 0.80 (0.76, 0.83) for weeks 8 and 12 for participants with OSA (Figure 2).

3.5 | Test–retest reliability of the ESS scores in the 12-week trials (by indication and treatment)

In the 12-week trial in participants with narcolepsy, the ICC estimates (95% CI) ranged from 0.81 (0.69, 0.89) for weeks 4 and 8 to

0.86 (0.77, 0.92) for weeks 8 and 12 among participants who received placebo and from 0.79 (0.73, 0.85) for weeks 4 and 12 to 0.86 (0.81, 0.90) for weeks 8 and 12 among participants who received solriamfetol. In the 12-week trial in participants with OSA, these values ranged from 0.62 (0.49, 0.73) for weeks 4 and 8 to 0.77 (0.68, 0.84) for weeks 8 and 12 among participants who received placebo and from 0.73 (0.67, 0.77) for weeks 4 and 12 to 0.79 (0.74, 0.83) for weeks 8 and 12 among participants who received solriamfetol (Figure 3).

3.6 | Test–retest reliability of the ESS scores in the OSA populations in the 12-week and 1-year trials (by adherence/non-adherence to primary OSA therapy)

In the 12-week trial, the ICC estimates (95% CI) ranged from 0.72 (0.66, 0.77) for weeks 4 and 12 to 0.80 (0.75, 0.84) for weeks 8 and 12 among participants who were adherent and from 0.73 (0.63, 0.80) for weeks 4 and 8 to 0.81 (0.74, 0.86) for weeks 8 and 12 among participants who were non-adherent to primary OSA therapy. In the OLE trial, the ICC estimates (95% CI) ranged from 0.74 (0.68, 0.79) for weeks 14 and 39/40 to 0.78 (0.72, 0.82) for weeks 26/27 and 39/40 among participants who were adherent and from 0.72 (0.58, 0.82) for weeks 14 and 39/40 to 0.78 (0.67, 0.86) for weeks 14 and 26/27 among participants who were non-adherent (Figure 4).
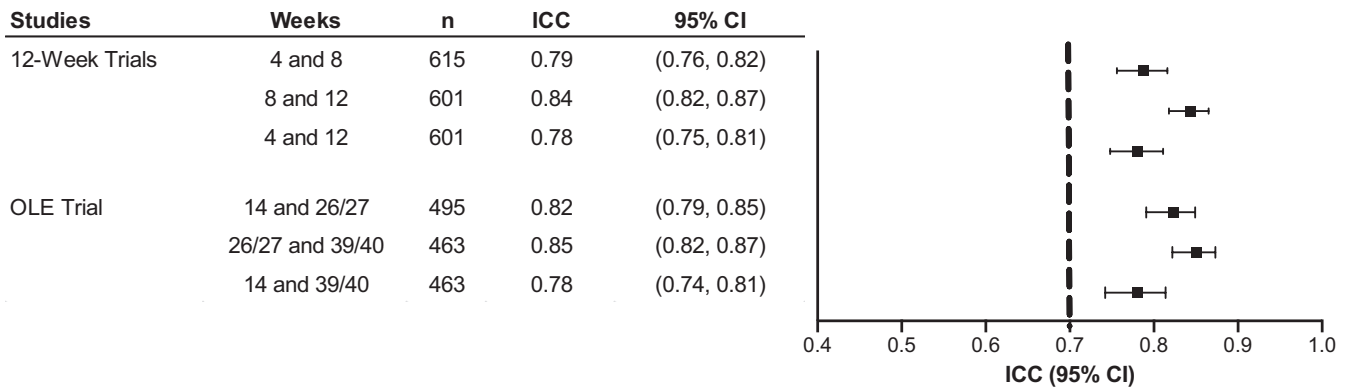
Studies	Weeks	n	ICC	95% CI
12-Week Trials	4 and 8	615	0.79	(0.76, 0.82)
	8 and 12	601	0.84	(0.82, 0.87)
	4 and 12	601	0.78	(0.75, 0.81)
OLE Trial	14 and 26/27	495	0.82	(0.79, 0.85)
	26/27 and 39/40	463	0.85	(0.82, 0.87)
	14 and 39/40	463	0.78	(0.74, 0.81)

ICC (95% CI)

FIGURE 1 Test–retest reliability of ESS scores in solriamfetol 12-week narcolepsy and OSA trials (pooled data) and long-term OLE trial. Dashed line represents recommended threshold for acceptable test–retest reliability (ICC point estimates >0.7) (Terwee et al., 2007). CI, confidence interval; ESS, Epworth Sleepiness Scale; ICC, intraclass correlation coefficient; OLE, open-label extension; OSA, obstructive sleep apnea
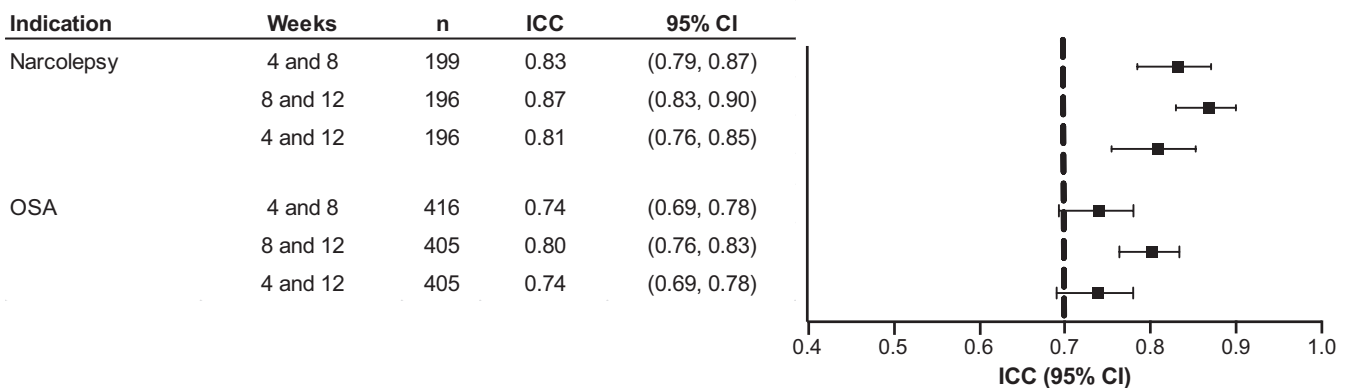
Indication	Weeks	n	ICC	95% CI
Narcolepsy	4 and 8	199	0.83	(0.79, 0.87)
	8 and 12	196	0.87	(0.83, 0.90)
	4 and 12	196	0.81	(0.76, 0.85)
OSA	4 and 8	416	0.74	(0.69, 0.78)
	8 and 12	405	0.80	(0.76, 0.83)
	4 and 12	405	0.74	(0.69, 0.78)

ICC (95% CI)

FIGURE 2 Test–retest reliability of ESS scores in solriamfetol 12-week trials by indication (narcolepsy or OSA). Dashed line represents recommended threshold for acceptable test–retest reliability (ICC point estimates >0.7) (Terwee et al., 2007). CI, confidence interval; ESS, Epworth Sleepiness Scale; ICC, intraclass correlation coefficient; OSA, obstructive sleep apnea

4 | DISCUSSION

In the present analysis, the ESS consistently demonstrated acceptable test–retest reliability across three large clinical trials of varying durations in clinical populations of participants with EDS associated with narcolepsy or OSA, regardless of treatment with placebo or a wake-promoting agent (solriamfetol) and, in participants with OSA, independent of level of adherence or non-adherence to primary OSA therapy. Specifically, the majority of ICC point estimates were >0.7, a threshold that has been recommended as a quality criterion for acceptable test–retest reliability (Terwee et al., 2007). The ICC point estimates below the threshold of <0.7 were observed only for placebo-treated participants with OSA for weeks 4 and 8 (0.62) and weeks 4 and 12 (0.67); lower bound 95% CI values <0.7 were also observed in comparisons indicating reliability could potentially fall below this threshold at times. These findings support use of the ESS as a reliable measure of EDS in clinical trials.

In contrast with the present findings, some studies have reported different measures of variability in ESS scores among patients with diagnosed or suspected sleep disorders outside the clinical trial setting (e.g. in clinical practice) (Campbell et al., 2018; Lee et al., 2020; Nguyen et al., 2006; Rozgonyi et al., 2021; Taylor et al., 2019; Walker et al., 2020). Nguyen et al. (Nguyen et al., 2006) found that 41% of patients had an ESS score difference of ≥3 points between the first and second assessment, 23% had a difference ≥5, and 10% had a difference ≥7. Similarly, Campbell et al. (Campbell et al., 2018) found that 46%, 21%, and 8% of patients had differences ≥3, ≥5, and ≥7 points, respectively. Taylor et al. (Taylor et al., 2019) found that 46%–72% of patients had differences of ≥2 points and 3%–21% had differences of ≥8 points. Lee et al. (Lee et al., 2020) found that 56% of sleep clinic patients had a difference of ≥3 points; however, the ICC was 0.73 (95% CI 0.61, 0.82). Rozgonyi et al. (Rozgonyi et al., 2021) found that in patients referred to a sleep clinic (who may or may not have had a diagnosis of a sleep disorder), Lin's concordance coefficient

(a) Narcolepsy

Treatment	Weeks	n	ICC	95% CI
Placebo	4 and 8	53	0.81	(0.69, 0.89)
	8 and 12	52	0.86	(0.77, 0.92)
	4 and 12	52	0.82	(0.70, 0.89)
Solriamfetol	4 and 8	146	0.83	(0.77, 0.87)
	8 and 12	144	0.86	(0.81, 0.90)
	4 and 12	144	0.79	(0.73, 0.85)
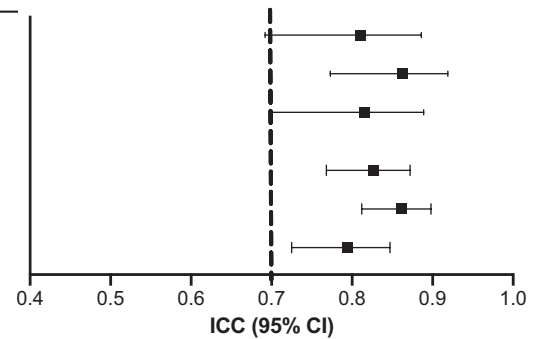
(b) OSA

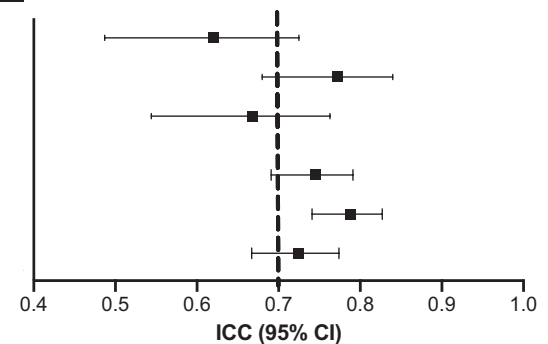| Treatment	Weeks	n	ICC	95% CI
Placebo	4 and 8	105	0.62	(0.49, 0.73)
	8 and 12	102	0.77	(0.68, 0.84)
	4 and 12	102	0.67	(0.54, 0.76)
Solriamfetol	4 and 8	311	0.75	(0.69, 0.79)
	8 and 12	303	0.79	(0.74, 0.83)
	4 and 12	303	0.73	(0.67, 0.77)

FIGURE 3 Test–retest reliability of ESS scores in solriamfetol 12-week trials by indication (narcolepsy or OSA) and treatment (placebo or combined solriamfetol). Dashed line represents recommended threshold for acceptable test-retest reliability (ICC point estimates >0.7) (Terwee et al., 2007). CI, confidence interval; ESS, Epworth Sleepiness Scale; ICC, intraclass correlation coefficient; OSA, obstructive sleep apnea

was 0.75 (scores of <0.9 indicate poor reliability) for pairs of ESS assessments that were an hour apart. Walker et al. (Walker et al., 2020) also found variability in individual ESS scores but found substantial agreement when the ESS was analysed in a binary fashion (i.e. sleepy or normal) using a ESS score cut-off of ≥11; 89% of patients with ESS scores of ≥11 at the first assessment also had ESS scores of ≥11 at the second assessment (up to 90 days later). Several factors may account for the discrepancy in these findings. First, the methods for assessing test–retest reliability differed among studies. Notably, only one (Lee et al., 2020) used the ICC method, which was selected for use in the present study because it is an established statistical method for evaluation of test–retest reliability. Other methods, such as a naïve correlation analysis, are not sensitive to systematic differences in repeated measures. The present analysis evaluated data from prospective clinical trials, whereas most other studies retrospectively analysed data from chart reviews in clinical practice (Campbell et al., 2018; Lee et al., 2020; Nguyen et al., 2006; Taylor et al., 2019; Walker et al., 2020). Further, in the retrospective chart review analyses, there was no control of, or means of assessing, other factors that may have changed between the first and second assessments and impacted intra-participant variability of EDS (e.g. change in total sleep time,

medication, or caffeine use). Finally, there was variability in how the test was administered or completed. In many cases, the first and second ESS assessments were administered in different settings (primary care versus specialist) (Campbell et al., 2018; Lee et al., 2020; Nguyen et al., 2006; Taylor et al., 2019; Walker et al., 2020). Indeed, Taylor et al. (Taylor et al., 2019) found low ICC estimates (0.31–0.34) for assessments that were administered in different clinical settings; however, when both assessments occurred in the same setting, the ICC estimate was much higher at 0.82.

The present analysis also found that the ESS has acceptable test–retest reliability in participants with narcolepsy. This finding is consistent with a previous study that analysed data from a randomised, controlled clinical trial of pitolisant, modafinil, or placebo (van der Heide et al., 2015). Specifically, van der Heide et al. (van der Heide et al., 2015) found the ICC estimate for the ESS to be 0.83 among participants with narcolepsy. These consistent findings are likely attributable to the fact that both analyses were based on data from clinical trials, in which assessment settings were similar across all time-points and factors that could affect variability of EDS (e.g. use of caffeine or other medications) were controlled.

A strength of the present analysis is the use of the ICC, an established statistical metric for evaluation of test–retest reliability. In

(a) 12-Week Trial

Adherence	Weeks	n	ICC	95% CI
Adherent	4 and 8	295	0.75	(0.69, 0.79)
	8 and 12	286	0.80	(0.75, 0.84)
	4 and 12	286	0.72	(0.66, 0.77)
Nonadherent	4 and 8	121	0.73	(0.63, 0.80)
	8 and 12	119	0.81	(0.74, 0.86)
	4 and 12	119	0.79	(0.71, 0.85)
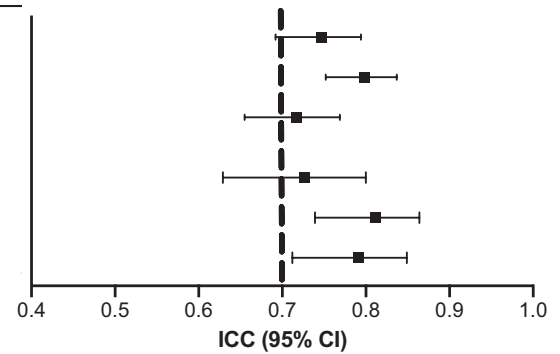
ICC (95% CI)

(b) OLE Trial

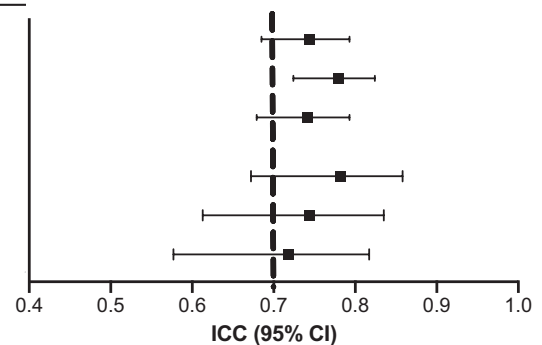| Adherence	Weeks	n	ICC	95% CI
Adherent	14 and 26/27	264	0.74	(0.69, 0.79)
	26/27 and 39/40	245	0.78	(0.72, 0.82)
	14 and 39/40	245	0.74	(0.68, 0.79)
Nonadherent	14 and 26/27	72	0.78	(0.67, 0.86)
	26/27 and 39/40	66	0.74	(0.61, 0.84)
	14 and 39/40	66	0.72	(0.58, 0.82)

ICC (95% CI)

FIGURE 4 Test–retest reliability of ESS scores in solriamfetol 12-week and long-term OLE trials in participants with OSA by adherence/non-adherence to primary OSA therapy. Dashed line represents recommended threshold for acceptable test–retest reliability (ICC point estimates >0.7) (Terwee et al., 2007). CI, confidence interval; ESS, Epworth Sleepiness Scale; ICC, intraclass correlation coefficient; OLE, open-label extension; OSA, obstructive sleep apnea

addition, reliability was assessed in a large data set of participants with EDS-associated sleep disorders across a variety of time intervals. Further, potential confounding factors that could contribute to intra-participant variability, such as the setting in which the ESS was administered and medication/caffeine use, were controlled. Finally, the present analysis included an exploratory evaluation of test–retest reliability in subgroups defined by treatment status (placebo or a wake-promoting agent) and adherence to primary OSA therapy. Despite these strengths, these clinical trials were not specifically designed to assess test–retest reliability of the ESS. Another limitation is that the generalisability of these findings to populations beyond patients with OSA or narcolepsy or to settings other than clinical trials is unknown.

In conclusion, the ESS consistently demonstrated acceptable test–retest reliability in clinical trial settings of OSA and narcolepsy across several months. Reliability remained acceptable across clinical trials of two different durations and in two different patient populations. Further, reliability was robust across subgroups defined by treatment (placebo or a wake-promoting agent) and by adherence/non-adherence to primary OSA therapy. These data from clinical trials suggest that the ESS is a reliable measure to assess the effects of treatment intervention on EDS in well-conducted clinical trial settings.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the study. DM conducted the statistical analyses. All authors contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content, and approved the final version of the manuscript.

CONFLICT OF INTEREST

RR has received consultancy fees from Eisai; honoraria from Merck; and research funding from Jazz Pharmaceuticals, Merck, Actelion, Eisai, and Philips Respironics; and has served on the speakers' bureau for Merck and as a board member for Jazz Pharmaceuticals. KB, DM, and MB are employees of Jazz Pharmaceuticals who, in the course of their employment, have received stock options exercisable for, and other stock awards of, ordinary shares of Jazz Pharmaceuticals plc. DH and SM are former employees of Jazz

Pharmaceuticals who, in the course of their employment, received stock options exercisable for, and other stock awards of, ordinary shares of Jazz Pharmaceuticals plc. JB is a part-time employee of Jazz Pharmaceuticals and shareholder of Jazz Pharmaceuticals plc.

ORCID
Russell Rosenberg https://orcid.org/0000-0002-6463-0082

REFERENCES

Ahmed, A. E., Fatani, A., Al-Harbi, A., Al-Shimemeri, A., Ali, Y. Z., Baharoon, S., & Al-Jahdali, H. (2014). Validation of the Arabic version of the Epworth Sleepiness Scale. *Journal of Epidemiology and Global Health*, 4(4), 297–302. https://doi.org/10.1016/j.jegh.2014.04.004

Black, J. E., & Hirshkowitz, M. (2005). Modafinil for treatment of residual excessive sleepiness in nasal continuous positive airway pressure-treated obstructive sleep apnea/hypopnea syndrome. *Sleep*, 28(4), 464–471. https://doi.org/10.1093/sleep/28.4.464

Black, J. E., Hull, S. G., Tiller, J., Yang, R., & Harsh, J. R. (2010). The long-term tolerability and efficacy of armodafinil in patients with excessive sleepiness associated with treated obstructive sleep apnea, shift work disorder, or narcolepsy: An open-label extension study. *Journal of Clinical Sleep Medicine*, 6(5), 458–466. https://doi.org/10.5664/jcsm.27935

Bogan, R. K., Feldman, N., Emsellem, H. A., Rosenberg, R., Lu, Y., Bream, G., Khayrallah, M., & Lankford, D. A. (2015). Effect of oral JZP-110 (ADX-N05) treatment on wakefulness and sleepiness in adults with narcolepsy. *Sleep Medicine*, 16(9), 1102–1108. https://doi.org/10.1016/j.sleep.2015.05.013

Broderick, J. E., Junghaenel, D. U., Schneider, S., Pilosi, J. J., & Stone, A. A. (2013). Pittsburgh and Epworth sleep scale items: accuracy of ratings across different reporting periods. *Behavioral Sleep Medicine*, 11(3), 173–188. https://doi.org/10.1080/15402002.2012.654549

Campbell, A. J., Neill, A. M., & Scott, D. A. R. (2018). Clinical reproducibility of the Epworth Sleepiness Scale for patients with suspected sleep apnea. *Journal of Clinical Sleep Medicine*, 14(5), 791–795. https://doi.org/10.5664/jcsm.7108

Crook, S., Sievi, N. A., Bloch, K. E., Stradling, J. R., Frei, A., Puhan, M. A., & Kohler, M. (2019). Minimum important difference of the Epworth Sleepiness Scale in obstructive sleep apnoea: estimation from three randomised controlled trials. *Thorax*, 74(4), 390–396. https://doi.org/10.1136/thoraxjnl-2018-211959

Dauvilliers, Y., Bassetti, C., Lammers, G. J., Arnulf, I., Mayer, G., Rodenbeck, A., ... Schwartz, J.-C. (2013). Pitolisant versus placebo or modafinil in patients with narcolepsy: a double-blind, randomised trial. *Lancet Neurology*, 12(11), 1068–1075. https://doi.org/10.1016/s1474-4422(13)70225-4

Dauvilliers, Y., Verbraecken, J., Partinen, M., Hedner, J., Saaresranta, T., Georgiev, O., ... Pepin, J. L. (2020). Pitolisant for daytime sleepiness in obstructive sleep apnea patients refusing CPAP: a randomized trial. *American Journal of Respiratory and Critical Care Medicine*, 201(9), 1135–1145. https://doi.org/10.1164/rccm.201907-1284OC

Harsh, J. R., Hayduk, R., Rosenberg, R., Wesnes, K. A., Walsh, J. K., Arora, S., ... Roth, T. (2006). The efficacy and safety of armodafinil as treatment for adults with excessive sleepiness associated with narcolepsy. *Current Medical Research and Opinion*, 22(4), 761–774. https://doi.org/10.1185/030079906x100050

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. *Sleep*, 14(6), 540–545. https://doi.org/10.1093/sleep/14.6.540

Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, 15(4), 376–381. https://doi.org/10.1093/sleep/15.4.376

Johns, M., & Hocking, B. (1997). Daytime sleepiness and sleep habits of Australian workers. *Sleep*, 20(10), 844–849. https://doi.org/10.1093/sleep/20.10.844

Kendzerska, T. B., Smith, P. M., Brignardello-Petersen, R., Leung, R. S., & Tomlinson, G. A. (2014). Evaluation of the measurement properties of the Epworth Sleepiness Scale: a systematic review. *Sleep Medicine Reviews*, 18(4), 321–331. https://doi.org/10.1016/j.smrv.2013.08.002

Knutson, K. L., Rathouz, P. J., Yan, L. L., Liu, K., & Lauderdale, D. S. (2006). Stability of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Questionnaires over 1 year in early middle-aged adults: the CARDIA study. *Sleep*, 29(11), 1503–1506. https://doi.org/10.1093/sleep/29.11.1503

Lee, J. L., Chung, Y., Waters, E., & Vedam, H. (2020). The Epworth sleepiness scale: Reliably unreliable in a sleep clinic population. *Journal of Sleep Research*, 29(5), e13019. https://doi.org/10.1111/jsr.13019

Malhotra, A., Shapiro, C., Pepin, J.-L., Hedner, J., Ahmed, M., Foldvary-Schaefer, N., ... Schwab, R. (2020). Long-term study of the safety and maintenance of efficacy of solriamfetol (JZP-110) in the treatment of excessive sleepiness in participants with narcolepsy or obstructive sleep apnea. *Sleep*, 43(2), zsz220. https://doi.org/10.1093/sleep/zsz220

Nguyen, A. T., Baltzan, M. A., Small, D., Wolkove, N., Guillon, S., & Palayew, M. (2006). Clinical reproducibility of the Epworth Sleepiness Scale. *Journal of Clinical Sleep Medicine*, 2(2), 170–174. https://doi.org/10.5664/jcsm.26512

Patel, S., Kon, S. S. C., Nolan, C. M., Barker, R. E., Simonds, A. K., Morrell, M. J., & Man, W. D. (2018). The Epworth Sleepiness Scale: Minimum clinically important difference in obstructive sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 197(7), 961–963. https://doi.org/10.1164/rccm.201704-0672LE

Plazzi, G., Ruoff, C., Lecendreux, M., Dauvilliers, Y., Rosen, C. L., Black, J., ... Mignot, E. (2018). Treatment of paediatric narcolepsy with sodium oxybate: a double-blind, placebo-controlled, randomised-withdrawal multicentre study and open-label investigation. *The Lancet Child & Adolescent Health*, 2(7), 483–494. https://doi.org/10.1016/s2352-4642(18)30133-0

Rozgonyi, R., Dombi, I., Janszky, J., Kovács, N., & Faludi, B. (2021). Low test-retest reliability of the Epworth Sleepiness Scale within a substantial short time frame. *Journal of Sleep Research*, 30(4), e13277. https://doi.org/10.1111/jsr.13277

Ruoff, C., Swick, T. J., Doekel, R., Emsellem, H. A., Feldman, N. T., Rosenberg, R., ... Black, J. (2016). Effect of oral JZP-110 (ADX-N05) on wakefulness and sleepiness in adults with narcolepsy: a phase 2b study. *Sleep*, 39(7), 1379–1387. https://doi.org/10.5665/sleep.5968

Schweitzer, P. K., Rosenberg, R., Zammit, G. K., Gotfried, M., Chen, D., Carter, L. P., ... Strohl, K. P. (2019). Solriamfetol for excessive sleepiness in obstructive sleep apnea (TONES 3): A randomized controlled trial. *American Journal of Respiratory and Critical Care Medicine*, 199(11), 1421–1431. https://doi.org/10.1164/rccm.201806-1100OC

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. https://doi.org/10.1037//0033-2909.86.2.420

Strollo, P. J., Hedner, J., Collop, N., Lorch, D. G., Chen, D., Carter, L. P., Lu, Y., ... Tones 4 Study Investigators (2019). Solriamfetol for the treatment of excessive sleepiness in OSA: A placebo-controlled randomized withdrawal study. *Chest*, 155(2), 364–374. https://doi.org/10.1016/j.chest.2018.11.005

Sunosi™ (solriamfetol) tablets Prescribing Information (2019). Jazz Pharmaceuticals Inc., Palo Alto, CA, USA. Retrieved from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/211230s000lbl.pdf

Sunosi™ (solriamfetol) tablets Summary of Product Characteristics (2020). Jazz Pharmaceuticals Ireland Ltd., Dublin, Ireland. Retrieved from https://www.ema.europa.eu/en/medicines/human/EPAR/sunosi

Szakacs, Z., Dauvilliers, Y., Mikhaylov, V., Poverennova, I., Krylov, S., Jankovic, S., Sonka, K., Lehert, P., Lecomte, I., Lecomte, J.-M., & Schwartz, J.-C., … for the HARMONY-CTP study group (2017). Safety and efficacy of pitolisant on cataplexy in patients with narcolepsy: A randomised, double-blind, placebo-controlled trial. *Lancet Neurology*, *16*(3), 200–207. https://doi.org/10.1016/s1474-4422(16)30333-7

Taylor, E., Zeng, I., & O'Dochartaigh, C. (2019). The reliability of the Epworth Sleepiness Score in a sleep clinic population. *Journal of Sleep Research*, *28*(2), e12687. https://doi.org/10.1111/jsr.12687

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Thorpy, M. J., Shapiro, C., Mayer, G., Corser, B. C., Emsellem, H., Plazzi, G., … Dauvilliers, Y. (2019). A randomized study of solriamfetol for excessive sleepiness in narcolepsy. *Annals of Neurology*, *85*(3), 359–370. https://doi.org/10.1002/ana.25423

US Department of Health and Human Services (2009). *Guidance for industry: Patient-reported outcome measures: use in medical product development to support labeling claims*. Retrieved from http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf

US Modafinil in Narcolepsy Multicenter Study Group (2000). Randomized trial of modafinil as a treatment for the excessive daytime somnolence of narcolepsy. *Neurology*, *54*(5), 1166–1175.

van der Heide, A., van Schie, M. K. M., Lammers, G. J., Dauvilliers, Y., Arnulf, I., Mayer, G., … van Dijk, J. G. (2015). Comparing treatment effect measurements in narcolepsy: The Sustained Attention to Response Task, Epworth Sleepiness Scale and Maintenance of Wakefulness Test. *Sleep*, *38*(7), 1051–1058. https://doi.org/10.5665/sleep.4810

Walker, N. A., Sunderram, J., Zhang, P., Lu, S. E., & Scharf, M. T. (2020). Clinical utility of the Epworth sleepiness scale. *Sleep & Breathing = Schlaf & Atmung*, *24*(4), 1759–1765. https://doi.org/10.1007/s11325-020-02015-2