SUPPLEMENTAL MATERIALS

Supplemental Methods

Treatment Phase: Structure of Analyses

During the treatment phase, 3 hypotheses were tested for the primary and key secondary endpoints. The first hypothesis was a validation test of the sensitivity and integrity of the study and tested whether the positive control (C) produced mean responses that show greater abuse potential compared with placebo (P). The first hypothesis type is expressed with these equations: $H_0: \mu_C - \mu_P \leq \delta_1$ versus $H_A: \mu_C - \mu_P > \delta_1$ where $\delta_1 > 0$. In all equations, H_0 is the null hypothesis, H_A is the alternative hypothesis, μ is the mean for each treatment group, and δ_1 is the prespecified validation margin. In this study, positive controls were zolpidem and suvorexant. For equations of this form, a significant *P* value, rejecting the null hypothesis, implies the difference between the positive control (zolpidem or suvorexant) and placebo exceeded the validation margin (15 or 11).

The validation margin (δ_1) was set to 15 for the primary endpoint, "at this moment" Drug Liking VAS. The validation margin of 15 was requested by the US Food and Drug Administration to match the margin of 15 used in the qualification phase. However, the originally planned validation margin was 11, which was based on determination of the clinically important difference on Drug Liking peak maximum effect in abuse potential studies[25-27] and was purposefully selected to be less stringent than the 15-point difference used for qualification purposes.[28] For the primary endpoint, results for each validation margin (15 and 11) are reported. For the key secondary endpoints, the validation margin (δ_1) was set to 11.[25] If the treatment difference $\mu_C - \mu_P$ was statistically significant and the lower confidence limit for the difference exceeded the validation margin (15 or 11), then validity was established for the study, and the other pairwise comparison tests were allowed.

The second hypothesis tested whether the test drug (T) produced mean responses that show less abuse potential compared with positive control and is expressed with these equations: $H_0: \mu_C - \mu_T \leq \delta_2$ versus $H_A: \mu_C - \mu_T > \delta_2$ where $\delta_2 \geq 0$. In this study, the test drug was lemborexant. δ_2 was set to 0 for the primary and key secondary endpoints. For equations of this form, a significant *P* value implies the positive control (zolpidem or suvorexant) and lemborexant was different (the difference between means was >0).

The third hypothesis tested whether the test drug produced mean responses that show similar abuse potential compared with placebo and is expressed with these equations: $H_0: \mu_T - \mu_P \geq \delta_3$ versus $H_A: \mu_T - \mu_P < \delta_3$ where $\delta_3 > 0$. The validation margin (δ_3) was set to 11 for the primary and key secondary endpoints. For equations of this form, a significant *P* value implies the difference between lemborexant and placebo was within the validation margin (11); that is, LEM was similar to PBO.

For endpoints that were not primary or key secondary, the study hypotheses took a standard form to test for differences between treatment groups. The equations were as follows: $H_0: \mu_C - \mu_P = 0$ versus $H_A: \mu_C - \mu_P \neq 0$; $H_0: \mu_C - \mu_T = 0$ versus $H_A: \mu_C - \mu_T \neq 0$; and $H_0: \mu_T - \mu_P = 0$ versus $H_A: \mu_T - \mu_P \neq 0$. For equations of this form, a significant *P* value implies the means in the 2 treatment groups were different.

Statistical Methods

As described in the Statistical Analyses subsection of the Methods, pharmacodynamic endpoints were analyzed using mixed-effect models, if the residuals from the model were normally distributed. The measures Drug Liking visual analog scale (VAS), Good Effects VAS, and Any Effects VAS were analyzed using the following model: least squares means were estimated from a mixed-effect model having treatment, period, and treatment sequence as fixed

effects, and subject nested within sequence as a random effect. Overall treatment effect was assessed using Friedman's test.

As necessary, first-order carryover effects and baseline (predose) measurements were included in the mixed-effect model. Observer's assessment of alertness/sedation (OAA/S) sum score; the choice reaction time (RT), recognition RT measure; and the divided attention task root mean square distance, greatest distance, percentage over road, and percentage of target hits measures were analyzed using the following model: least squares means were estimated from a mixed-effect model having treatment, period, and treatment sequence as fixed effects, baseline (predose) measurement as a covariate, and subject nested within sequence as a random effect. Overall treatment effect was assessed using Friedman's test. Stoned VAS and the Alertness/Drowsiness VAS, Addiction Research Center Inventory for the pentobarbital-chlorpromazine-alcohol group scale were analyzed using the following model: least squares means were estimated from a mixed-effect model having treatment, period, treatment sequence, and first-order carryover as fixed effects, baseline (predose) measurement as a covariate, and subject nested within sequence as a random effect.

If the residuals from the mixed model were not normally distributed, paired *t* tests were used to assess mean treatment differences, if the distribution of paired differences was normal or quite symmetric. If paired differences were not normally distributed, the Wilcoxon signed rank test was used to assess median treatment differences. Within each measure, pairwise between-treatment comparisons were conducted using the *t* test or signed rank test, as appropriate based on normality. Therefore, both *t* tests and signed rank tests were used in some cases for different between-treatment comparisons within the same outcome measure. The following measures were analyzed by pairwise comparison of means or medians: Overall Drug Liking VAS; Take Drug Again VAS; subjective drug value; Bad Effects VAS; Alertness/Drowsiness VAS; OAA/S composite score; the choice RT motor RT, total RT, and percentage correct

measures; and the divided attention task response latency of correct responses and number of false alarms measures.

TABLE S1. Description of Visual Analog Scales

Scale Interpretation	Exclude Predose	Description	Question Text	Response Anchors
Balance	Yes	Drug Liking	At this moment, my liking for this drug	0: Strong disliking 50: Neither like nor dislike 100: Strong liking
Global	Yes	Overall Drug Liking	Overall, my liking for this drug is	
Global	Yes	Take Drug Again	I would take this drug again	0: Definitely not 50: Neutral 100: Definitely so
Positive	Yes	Good Effects	At this moment, I feel good drug effects	0: Not at all 100: Extremely
Positive	No	High	At this moment, I feel high	
Positive	No	Stoned	At this moment, I feel stoned	
Negative	Yes	Bad Effects	At this moment, I feel bad drug effects	
Other effects	Yes	Any Effects	At this moment, I feel any drug effect	
Other effects	No	Alertness/Drowsiness	At this moment, my mental state is	0: Very drowsy 50: Neither drowsy nor alert 100: Very alert

Scales had 5 possible interpretations: (1) “Positive” subjective effects, (2) “Negative” subjective effects, (3) the “Balance” between positive and negative effects, (4) “Other effects” (ie, pharmacologic effects that indicate an active substance, which may be perceived as either positive or negative depending on the context), and (5) “Global” (ie, end-of-day assessment or next-day overall assessment).

TABLE S2. Findings for Primary Endpoint "At This Moment" Drug Liking VAS E_{max} (Completer Analysis Set)

Value	PBO (n = 32)	ZOL (n = 32)	SUV (n = 32)	LEM10 (n = 32)	LEM20 (n = 32)	LEM30 (n = 32)
Mean (SE)	57.8 (2.9)	78.3 (2.8)	76.1 (3.2)	78.4 (3.3)	80.5 (3.1)	83.6 (3.0)
LSM (95% CI)[a]	58.3 (52.3–64.3)	78.5 (72.5–84.5)	76.5 (70.5–82.5)	78.9 (72.9–84.9)	80.9 (74.9–86.9)	83.9 (77.9–89.9)
LSM (SE), Active – PBO		20.2 (3.7)[b,c,*†]	18.2 (3.7)[b,c,*†]	20.5 (3.7)[d,‡]	22.5 (3.7)[d,‡]	25.5 (3.7)[d,‡]
95% CI		Lower 95% CI: 14.1	Lower 95% CI: 12.2	Upper 95% CI: 26.6	Upper 95% CI: 28.6	Upper 95% CI: 31.6
LSM (SE), ZOL − LEM[e]				−0.3 (3.7)	−2.4 (3.7)	−5.4 (3.7)
LSM (SE), SUV − LEM[e]				−2.3 (3.7)	−4.3 (3.7)	−7.3 (3.7)

CI, confidence interval; E_{max}, maximum (peak) effect; LEM10, lemborexant 10 mg; LEM20, lemborexant 20 mg; LEM30, lemborexant 30 mg; LSM, least squares mean; PBO, placebo; SE, standard error; SUV, suvorexant 40 mg; VAS, visual analog scale; ZOL, zolpidem 30 mg.

*Indicates statistically significant difference for positive control (ZOL and SUV) versus PBO ($P < 0.05$).

†Indicates difference versus PBO was significant when assessed with a validation margin of 11, but not when assessed with a validation margin of 15.

‡Indicates statistically significant difference versus PBO (where $P > 0.05$ signifies treatments are significantly different).

[a]LSMs were estimated from a mixed-effect model having treatment, period, and treatment sequence as fixed effects, and subject nested within sequence as a random effect. Overall treatment effect was assessed using Friedman's test.

[b]For the assessment of study validity with a margin of 15, hypothesis tests were constructed as follows: H_0: $\mu_C - \mu_P \leq 15$ vs H_A: $\mu_C - \mu_P > 15$, where C = positive control (ZOL and SUV) and P = PBO.

[c]For the assessment of study validity with a margin of 11, hypothesis tests were constructed as follows: H_0: $\mu_C - \mu_P \leq 11$ vs H_A: $\mu_C - \mu_P > 11$, where C = positive control (ZOL and SUV) and P = PBO.

[d]For LEM versus PBO comparisons, hypothesis tests were constructed as follows: H_0: $\mu_T - \mu_P \geq 11$ vs. H_A: $\mu_T - \mu_P < 11$ where T = test drug (LEM) and P = PBO.

[e]For LEM versus positive control (ZOL and SUV) comparisons, hypothesis tests were constructed as follows: H_0: $\mu_C - \mu_T \leq 0$ vs. H_A: $\mu_C - \mu_T > 0$ where C = positive control (ZOL or SUV) and T = test drug (LEM).

TABLE S3. Findings for Secondary Endpoint Measures (Completer Analysis Set)

Measure, value	PBO (n = 32)	ZOL (n = 32)	SUV (n = 32)	LEM10 (n = 32)	LEM20 (n = 32)	LEM30 (n = 32)
Overall Drug Liking VAS, score, E_{max}						
Mean (SE)	54.7 (2.2)	75.6 (4.1)	79.0 (3.8)	76.6 (4.0)	78.2 (4.1)	77.3 (3.7)
Active – PBO		23.0 (5.5, 42.5)[a,*]	27.5 (3.0, 49.5)[a,*]	22.0 (4.1)[b,†]	29.5 (6.5, 49.0)[b,†]	27.5 (5.5, 47.5)[b,†]
ZOL – LEM[c]				−1.0 (4.7)	−2.6 (4.3)	−1.7 (4.2)
SUV – LEM[c]				2.4 (4.7)	0.8 (4.2)	1.8 (3.7)
Take Drug Again VAS, score, E_{max}						
Mean (SE)	55.5 (2.3)	78.7 (4.4)	79.3 (3.9)	78.2 (3.8)	79.8 (4.2)	78.2 (4.4)
Active – PBO		24.5 (8.5, 48.5)[a,*]	23.7 (4.3)[a,*]	22.7 (4.0)[b,†]	29.5 (8.0, 49.0)[b,†]	34.5 (−1.5, 49.0)[b,†]
ZOL – LEM[c]				0.4 (4.8)	−1.2 (4.8)	0.4 (4.3)
SUV – LEM[c]				1.0 (3.8)	0 (−7.0, 3.5)	1.0 (4.2)
Subjective drug value, $, E_{max}[d]						
Mean (SE)	2.65 (1.26)	16.55 (2.83)	13.74 (2.74)	14.44 (2.63)	16.92 (2.73)	14.88 (2.36)
Active – PBO		13.90 (3.20)*	7.75 (2.25, 12.50)*	8.75 (0, 20.75)*	14.27 (3.07)*	12.23 (2.82)*
ZOL – LEM				0.00 (−4.88, 6.13)	−0.38 (2.64)	1.67 (3.00)
SUV – LEM				0.00 (−8.00, 4.75)	−0.50 (−7.50, 2.75)	−0.38 (−7.13, 3.88)
Good Effects VAS, score, E_{max}						
Mean (SE)	13.6 (4.56)	69.2 (5.02)	50.9 (6.31)	64.3 (5.88)	71.5 (5.18)	77.8 (4.57)
Active – PBO		55.7 (6.35)*	37.3 (7.93)*	50.8 (6.49)†	57.9 (6.32)†	64.3 (5.91)†
ZOL – LEM				4.9 (7.03)	−2.3 (6.59)	−8.6 (6.25)
SUV – LEM				−13.4 (8.06)	−20.6 (6.42)	−26.9 (6.58)
Stoned VAS, score, E_{max}						

Mean (SE)	11.6 (4.26)	59.1 (6.15)	30.6 (6.31)	46.4 (6.93)	52.8 (6.15)	62.5 (6.79)
Active – PBO		47.5 (7.45)*	18.9 (6.97)*	34.8 (6.82)	41.2 (7.06)	50.8 (6.70)
ZOL – LEM				12.8 (8.15)	6.3 (7.98)	−3.3 (8.21)
SUV – LEM				−15.8 (7.42)*	−22.2 (7.12)*	−31.9 (6.47)*
High VAS, score, E_{max}						
Mean (SE)	14.4 (4.85)	65.6 (4.59)	39.3 (5.88)	59.9 (6.32)	65.5 (5.73)	82.7 (4.49)
Active – PBO		51.2 (6.62)*	24.9 (7.04)*	45.6 (7.00)	51.2 (7.47)	68.3 (6.47)
ZOL – LEM				5.6 (8.08)	0.0 (6.41)	-17.1 (5.98)
SUV – LEM				−20.6 (7.63)	−26.2 (6.14)	−43.3 (7.13)
Bad Effects VAS, score, E_{max}						
Mean (SE)	5.5 (2.45)	42.0 (6.10)	14.4 (4.31)	26.3 (5.69)	35.8 (6.94)	37.5 (6.53)
Active – PBO		36.5 (6.80)*	1.0 (0.0, 13.5)*	5.5 (0.0, 40.5)†	11.5 (0.0, 60.0)†	13.0 (2.0, 51.0)†
ZOL – LEM				15.7 (7.07)*	6.2 (7.49)	4.5 (7.25)
SUV – LEM				−11.9 (4.98)*	−2.5 (−54.5, 0.0)*	−8.5 (−50.5, 0.0)*
Alertness/Drowsiness VAS, score, E_{min}						
Mean (SE)	40.4 (2.74)	15.5 (2.32)	16.4 (2.53)	8.6 (2.08)	6.9 (1.64)	6.0 (1.83)
Active – PBO		−24.9 (3.47)*	−24.0 (3.26)*	−33.0 (−48.0, −22.5)†	−36.0 (−48.5, −24.0)†	−40.0 (−49.5, −23.0)†
ZOL – LEM				6.9 (2.68)*	2.0 (0.5, 16.5)*	10.0 (0.0, 22.5)*
SUV – LEM				7.8 (3.20)*	9.6 (2.91)*	12.0 (0.0, 20.5)*
ARCI (Pentobarbital-Chlorpromazine-Alcohol Group Scale) score, E_{max}						
Mean (SE)	5.2 (0.66)	10.9 (0.58)	9.6 (0.53)	10.5 (0.49)	11.5 (0.49)	11.3 (0.46)
Active – PBO		5.7 (0.60)*	4.4 (0.68)*	5.3 (0.62)†	6.4 (0.70)†	6.2 (0.63)†
ZOL – LEM				0.40 (0.40)	-0.7 (0.40)	-0.5 (0.42)

SUV – LEM				−0.9 (0.44)	−1.9 (0.49)†	−1.8 (0.41)†
Any Effects VAS, score, E_{max}						
Mean (SE)	13.3 (4.76)	74.6 (4.70)	56.0 (6.43)	76.5 (5.91)	83.6 (3.73)	89.0 (2.82)
Active – PBO		61.3 (6.64)*	42.7 (7.14)*	63.2 (6.54)†	70.3 (5.81)†	75.7 (4.95)†
ZOL – LEM				−1.9 (7.32)*	−9.1 (5.24)	−14.4 (5.07)
SUV – LEM				−20.5 (7.86)	−27.6 (5.68)*	−33.0 (5.83)*
Observer's Assessment of Alertness/Sedation, E_{min}						
Mean (SE)	4.4 (0.14)	2.3 (0.12)	3.3 (0.15)	3.1 (0.14)	2.9 (0.16)	2.8 (0.13)
Active – PBO		−2.2 (0.17)*	−1.0 (−2.0, −0.5)*	−1.3 (0.15)†	−1.0 (−2.0, −1.0)†	−1.6 (0.15)†
ZOL – LEM				−0.8 (0.15)	−1.0 (−1.0, 0.0)	−0.6 (0.17)*
SUV – LEM				0.0 (0.0, 1.0)*	0.4 (0.17)*	0.5 (0.17)*

For comparisons between treatments: If a paired *t* test was used to assess the difference between 2 treatments, mean (SE) difference is presented; if the sign test was used, median (1st and 3rd quartile) difference is presented.

*Indicates statistically significant difference versus comparator ($P < 0.05$).

†Indicates statistically significant difference versus PBO (where $P > 0.05$ signifies treatments are significantly different).

[a]For positive control (ZOL and SUV) versus PBO comparisons, hypothesis tests were constructed as follows: H_0: $\mu_C - \mu_P \leq 11$ vs H_A: $\mu_C - \mu_P > 11$, where C = positive control (ZOL and SUV) and P = PBO.

[b]For LEM versus PBO comparisons, hypothesis tests were constructed as follows: H_0: $\mu_T - \mu_P \geq 11$ vs. H_A: $\mu_T - \mu_P < 11$ where T = test drug (LEM) and P = PBO.

[c]For LEM versus positive control (ZOL and SUV) comparisons, hypothesis tests were constructed as follows: H_0: $\mu_C - \mu_T \leq 0$ vs. H_A: $\mu_C - \mu_T > 0$, where C = positive control (ZOL or SUV) and T = test drug (LEM).

[d]Comparisons tested the null hypothesis that the difference of the means between treatment groups is zero.

ARCI, Addition Research Center Inventory; E_{max}, maximum (peak) effect; E_{min}, minimum (peak) effect; LEM10, lemborexant 10 mg; LEM20, lemborexant 20 mg; LEM30, lemborexant 30 mg; PBO, placebo; SE, standard error; SUV, suvorexant 40 mg; VAS, visual analog scale; ZOL, zolpidem 30 mg.