

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**204569Orig1s000**

**STATISTICAL REVIEW(S)**



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Sciences  
Office of Biostatistics

## STATISTICAL REVIEW AND EVALUATION

### CLINICAL STUDIES

**NDA/BLA Serial**

**Number:** 204569 (0000)  
**Drug Name:** Suvorexant  
**Indication(s):** Insomnia  
**Applicant:** Merck  
**Date(s):** August 29, 2012

**Review Priority:** Standard

**Biometrics Division:** Division of Biometrics I

**Statistical Reviewer:** Tristan Massie, Ph.D.

**Concurring Reviewers:** Kun Jin, Ph.D., Team Leader  
Kooros Mahjoob, Ph.D. Deputy Director, Division of Biometrics I

**Medical Division:** Neurology (HFD-120)

**Clinical Team:** Kachikwu Illoh, M.D., Clinical Reviewer  
Ron Farkas, M.D. Team Leader  
Russell Katz, M.D., Division Director

**Project Manager:** Cathleen Michaloski

**Keyword(s):** Multiplicity; Dose-Response; Randomization Ratio; Dosing by Age Group

## Table of Contents

<b>STATISTICAL REVIEW AND EVALUATION .....</b>	<b>1</b>
<b>LIST OF TABLES.....</b>	<b>3</b>
<b>LIST OF FIGURES.....</b>	<b>4</b>
<b>1 EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>2 INTRODUCTION .....</b>	<b>6</b>
2.1 OVERVIEW.....	6
2.2 DATA SOURCES .....	7
<b>3 STATISTICAL EVALUATION .....</b>	<b>8</b>
3.1 EVALUATION OF EFFICACY .....	8
3.1.1 <i>Study P028</i> .....	8
3.1.1.1 Study Design and Statistical Methods.....	8
3.1.1.2 Patient Disposition .....	15
3.1.1.3 Baseline Demographics and Disease Characteristics .....	16
3.1.1.4 Sponsor’s Results.....	18
3.1.1.5 Reviewer’s Results.....	25
3.1.2 <i>Study P029</i> .....	29
3.2.1.1 Study Design and Statistical Methods.....	29
3.2.1.2 Patient Disposition .....	29
3.2.1.1 Patient Demographics and Baseline Disease Characteristics .....	30
3.1.2.1 Sponsor’s Results.....	32
3.1.2.2 Reviewer’s Results.....	39
3.1.3 <i>Study P006</i> .....	42
3.1.3.1 SUMMARY OF STUDY DESIGN .....	42
3.1.3.2 Patient Disposition .....	45
3.1.3.3 Patient Demographics .....	45
3.1.3.4 Sponsor’s Results.....	47
3.1.3.5 Reviewer’s Results.....	59
3.2 EVALUATION OF SAFETY .....	62
3.2.1 <i>Multiple Dose Studies to Evaluate Next Day Effects of MK-4305 (Suvorexant) on Driving Performance in Healthy Subjects</i> .....	62
<b>4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>65</b>
4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION .....	65
4.1.1 <i>Gender</i> .....	65
4.1.2 <i>Race</i> .....	69
4.1.3 <i>Age</i> .....	73
4.1.4 <i>Region</i> .....	76
4.1.5 <i>Site Effects</i> .....	80
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS .....	83
4.2.1 <i>Cohort</i> .....	83
<b>5 SUMMARY AND CONCLUSIONS .....</b>	<b>84</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	84
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	90

## LIST OF TABLES

Table 1 Key Suvorexant Efficacy Studies .....	7
Table 2 Study 28: Patient Disposition .....	15
Table 3 Study 28: Baseline Demographics (Reviewer’s Analysis) .....	17
Table 4 Study 28: Baseline Efficacy Measures (Reviewer’s Analysis).....	18
Table 5 Study 28: Mean Subjective Total Sleep Time (Sponsor’s Analysis).....	19
Table 6 Study 28: Mean subjective Time to Sleep Onset (Sponsor’s Analysis).....	21
Table 7 Study 28: Mean Objective WASO (Sponsor’s Analysis) .....	22
Table 8 Study 28 Mean Change from Baseline in Objective LPS (Sponsor’s Analysis) .....	23
Table 9 Study 29: Disposition of Patients .....	30
Table 10 Study 29: Baseline Demographics and Disease Characteristics (Randomized Patients).....	31
Table 11 Study 29: Change from Baseline in Mean Subjective Total Sleep Time (Sponsor’s Analysis) .....	34
Table 12 Study 29: Change from Baseline in Mean Objective WASO (Sponsor’s Analysis).....	35
Table 13 Study 29 Mean Change from Baseline in sTSO (Sponsor’s Analysis).....	36
Table 14 Study 29: Mean Change from Baseline in LPS (Sponsor’s Analysis).....	37
Table 15 Study 29: Primary Efficacy Results for High Dose Including Site 45490.....	41
Table 16 Study 29: Sensitivity Analysis for Primary Subjective Endpoints (High Dose vs. Placebo).....	41
Table 17 Study 06: Baseline Demographics by Sequence.....	46
Table 18 Study 06: Baseline Disease Characteristics by Sequence .....	47
Table 19 Study 06: P-value Summary of Sponsor’s Primary and Key Secondary Analyses.....	48
Table 20 Study 06: Sponsor’s Analysis of WASO Based on Both Periods.....	50
Table 21 Study 06 Sponsor’s exploratory Analysis of WASO for first period only.....	51
Table 22 Study 06: Sponsor’s Analysis of Latency to Persistent Sleep Based on Both Periods .....	52
Table 23 Study 06: Sponsor’s Analysis of Latency to Persistent Sleep Based on First Period Only .....	54
Table 24 Study 06: Summary Statistics for Latency to Persistent Sleep in First Period .....	55
Table 25 Study 06: Sponsor’s Analysis of Subjective TSO: Differences from Placebo .....	56
Table 26 Study 06: Analysis of sTSO in first period only.....	57
Table 27 Study 06: Sponsor’s Analysis of Subjective TST .....	58
Table 28 Study 06: Analysis of sTST in first period only .....	58
Table 29 Study 06: Reviewer’s Multiple Imputation Sensitivity Analyses for Missing Data .....	62
Table 30 Study 35: Non-Elderly Driving Study Analysis of Mean SDLP .....	63
Table 31 Study 35: Non-Elderly Driving Study Symmetry Analysis.....	64
Table 32 Study 39: Elderly Driving Study Analysis of Mean SDLP .....	64
Table 33 Study 39: Elderly Driving Study Symmetry Analysis .....	65
Table 34 Change from Baseline in LPS by Gender in P28/P29 Pooled .....	66
Table 35 Change from Baseline in WASO by Gender in P28/P29 Pooled.....	67
Table 36 Change from Baseline in sTSO by Gender in P28/P29 Pooled .....	67
Table 37 Change from Baseline in sTST by Gender in P28/P29 Pooled.....	68
Table 38 Study 06 First Period Only Analyses of Change from Baseline in WASO and LPS by Gender.....	68
Table 39 Study 28 Differences in Change from Baseline in sTSO High Dose vs. Placebo at Month 1 by Race .....	69
Table 40 Study 28: Differences in Change from Baseline in sTST at Month 1 High Dose vs. Placebo by Race .....	70
Table 41 Study 28: Differences in Change from Baseline in LPS for High Dose vs. Placebo by Race .....	70
Table 42 Study 28: Differences in Change from Baseline in WASO for High Dose vs. Placebo by Race .....	71
Table 43 Study 29: sTSO Differences for High Dose vs Placebo by Race .....	71
Table 44 Study 29: sTST Differences for High Dose vs Placebo by Race.....	72
Table 45 Study 29: WASO Differences for High Dose vs Placebo by Race.....	72
Table 46 Study 29: LPS Differences for High Dose vs. Placebo by Race.....	73
Table 47 Change from Baseline in LPS by Age Group in P28/P29 Pooled .....	74
Table 48 Change from Baseline in WASO by Age Group in P28/P29 Pooled Studies.....	74
Table 49 Change from Baseline in sTST by Age Group in P28/P29 Pooled.....	75
Table 50 Change from Baseline in sTSO by Age Group in P28/P29 Pooled.....	75
Table 51 Study 29: Regional Subgroup Analyses for LPS .....	76

Table 52 Study 29 Regional Subgroup Analyses for WASO .....	76
Table 53 Study 28 Regional Subgroup Analyses for LPS .....	77
Table 54 Study 28 Regional Subgroup Analyses for WASO .....	77
Table 55 Study 28 Regional Subgroup Analyses for sTST .....	78
Table 56 Study 28 Regional Subgroup Analyses for sTSO .....	78
Table 57 Study 29 Regional Subgroup Analyses for sTST .....	79
Table 58 Study 29 Regional Subgroup Analyses for sTSO .....	79
Table 59 Study 28: Cohort Differences for Subjective Endpoints for High Dose vs. Placebo .....	84
Table 60 Study 29: Cohort Differences for Subjective Endpoints for High Dose vs. Placebo .....	84
Table 61 Summary Statistics for Efficacy Measures at Baseline by Trial in the Phase 2b/3 Trials .....	85
Table 62 Summary of Efficacy Measures for Lowest Doses in Phase 2B/3 studies.....	86

## LIST OF FIGURES

Figure 1 Multiplicity Adjustment Method used in Phase 3 Studies.....	14
Figure 2 Study 28: Mean Subjective Total Sleep Time (Sponsor's Analysis).....	20
Figure 3 Study 28 Mean Change from Baseline in Objective LPS (Sponsor's Analysis) .....	24
Figure 4 Study 28: Percentages of Randomized Patients with non-Missing LPS data over Time.....	25
Figure 5 Study 28: Percentages of Randomized Patients with non-Missing sTST data over Time .....	26
Figure 6 Study 28 Mean Change from Baseline in sTSO over Time by Completion Status .....	28
Figure 7 Study 28 Mean Change from Baseline in WASO over Time by Completion Status .....	29
Figure 8 Study 29: Mean Change from Baseline in sTST over Time (Sponsor's Analysis) .....	33
Figure 9 Study 29: Adjusted Means for Change from Baseline in Latency to Persistent Sleep .....	38
Figure 10 Study 29: Percentages of Randomized Patients with non-Missing LPS data over Time .....	40
Figure 11 Study 06: Mean Change from Baseline in LPS by Sequence and Period.....	60
Figure 12 Study 06: Mean Change from Baseline in LPS during first period for the placebo first sequences.....	61
Figure 13 Study 28: Differences of Placebo from High Dose in Mean Change in LPS at Month 1 by Site .....	81
Figure 14 Study 29: Differences of Placebo from High Dose in Mean Change in LPS at Month 1 by Site .....	82
Figure 15 Study 29: Differences of Placebo from High Dose in Mean Change in Subj. TSO at Month 1 by Site.....	83
Figure 16 Study 06: Dose Response in First Period for Various Efficacy Measures .....	88
Figure 17 Study 28/29 Pooled Estimates for Primary Efficacy Measures .....	89

## 1 EXECUTIVE SUMMARY

The clinical trial efficacy data provided in this application seems to clearly support the efficacy of Suvorexant for Sleep Maintenance. In the application there are two similarly designed 3 month placebo controlled phase 3 studies (study 28 and study 29) and one early phase 2B dose finding crossover study. The evidence for an effect on Sleep Onset was weaker than that for maintenance. In one study at the 3 month visit night the effect on latency as measured objectively by Polysomnography did not achieve statistical significance, thus failing to replicate the statistically significant effect demonstrated in the other study. However, the high dose effect at Month 1 was significant in both studies and there was replication of the effect on the corresponding subjective assessment, the patient reported weekly average of the Time to Sleep Onset, at Month 3 (as well as Month 1).

The Suvorexant phase 3 studies were ambitious in that they aimed to demonstrate effects for both elderly and non-elderly patients on both Sleep Maintenance and Onset, in terms of both an objective and a subjective assessment for each, in the same study. They also put more weight on the high dose (the low dose had 30 to 40% less patients by design) and the multiplicity adjustment method tested the high dose first, such that the low dose could only be tested if the high dose was first significant at multiple timepoints on both subjective and objective assessments. After taking into account the prespecified adjustments for multiple testing the low dose was only statistically significant for objective latency to persistent sleep in study 28 (not significant for objective latency in study 29 or for subjective time to sleep onset in study 28 or study 29). However, the low dose was statistically significant for objective and subjective primary endpoints for maintenance in both of these studies.

In the non-elderly next day driving study the 40 mg dose had significant asymmetry of differences from placebo with significantly more being positive and higher than the impairment threshold of interest (2.40) on both days 2 and 9. The low dose also was significant on day 2 but not on day 9. There was also some other evidence that the driving effect might be dose related which would suggest considering a lower dose if it was still efficacious.

In a prior phase 2B crossover study in non-elderly adults a lower dose, 10 mg, as well as a higher dose, 80 mg, were studied in addition to the adult doses used later in Phase 3. There was a suggestion of efficacy of 10 mg, particularly for wake time after sleep onset, based on this study data but there is no existing means of replication for the 10 mg dose and no 10 mg data for the elderly.

## 2 INTRODUCTION

### 2.1 Overview

The IND number associated with the development of this drug for this indication is IND 101,847. Suvorexant is a selective antagonist for orexin receptors OX<sub>1</sub>R and OX<sub>2</sub>R. In the two confirmatory efficacy trials of suvorexant (P028 and P029), efficacy and safety of suvorexant were evaluated in replicate core 3-month Treatment Phases. These trials were similarly designed as combined-age (with enrollment of both non-elderly and elderly adults) and combined-measure studies (with data collected for both objective and subjective efficacy measures). Based on the results of the Phase 2b 2 period crossover trial (P006) comparing each of 10 mg, 20 mg, 40 mg, and 80 mg separately with placebo, the Phase 3 dose of primary focus, to be confirmed in non-elderly patients (< 65 years), was 40 mg (referred to as "suvorexant high dose [HD]"). Based on evidence for slightly higher exposures in elderly patients in Phase 1 studies, suvorexant HD in elderly patients was 30 mg. Thus exposure levels for suvorexant HD were anticipated to be similar for non-elderly and elderly patients enabling the pooling of efficacy and safety data across age groups. A lower dose (LD) of 20 mg in non-elderly and 15 mg in elderly was also evaluated in these two trials, but with a smaller sample size than HD (the sponsor's intention was to pool the samples across the two studies for more precise estimation of LD effects). Selection of patients with insomnia in the confirmatory trials (P028 and P029) relied on objective polysomnograph (PSG) criteria for the PQ-Cohort (patients with both PSG and e-diary data) and subjective questionnaire (e-diary) criteria for the Q-Cohort (patients with e-diary data only). Table 1 summarizes the features of the key efficacy studies.

**Table 1 Key Suvorexant Efficacy Studies**

Study	# of Subjects per Arm	Follow-up Period	Completer N (%)	Primary Efficacy	Study Population
P28: Phase 3	<u>N</u> Placebo: 385 LD 255 HD 383	3 months	341 (88.6) 230 (90.6) 345 (90.1)	<b><u>Latency</u>    <u>Maintenance</u></b> <b>LPS        WASO</b> <b>sTSO       sTST</b> <b>at months 1 and 3</b> <b>PSG and diary based</b> <b>measures</b>	42% Elderly 34% North America 62%Fe male
P29 Phase 3	<u>N</u> Placebo: 389 LD 240 HD 392	3 months	330 (85.3%) 205 (85.4%) 346 (88.3%)	<b><u>Latency</u>    <u>Maintenance</u></b> <b>LPS        WASO</b> <b>sTSO       sTST</b> <b>at months 1 and 3</b> <b>PSG and diary based</b> <b>measures</b>	41% Elderly 48% North America 67%Fe male
P006: Phase 2b	2 Period Crossover Placebo: 10 mg 10 mg: Placebo Placebo: 20 mg 20 mg: Placebo Placebo: 40 mg 40 mg: Placebo Placebo: 80 mg 80 mg: Placebo	1 month for each Period plus 1 week washout	31(94%) 32(94%) 33(91%) 32(81%) 32(94%) 32(84%) 31(90%) 31(90%)	<b>Sleep Efficiency at Night 1 and 28</b> <b>Secondary Endpoints</b> <b>WASO and LPS</b>	Non- Elderly 87% North America 58% Female

**2.2 Data Sources**

At the time of review the locations of the primary endpoint data for the two key studies were as follows. The polysomnographic endpoint data are in the files ADPSG and the subjective endpoint weekly average data are in the files ADMD within these directories.

<\\Cdsub1\evsprod\NDA204569\0000\m5\datasets\p028\analysis\datasets>

<\\Cdsub1\evsprod\NDA204569\0000\m5\datasets\p029\analysis\datasets>

Analysis datasets for the phase 2b crossover study, P006, were not provided in the original submission but were provided later during the review following an FDA request for them.

<\\Cdsub1\evsprod\NDA204569\0014\m5\datasets\p006\analysis\datasets>

### 3 STATISTICAL EVALUATION

#### 3.1 Evaluation of Efficacy

##### 3.1.1 Study P028

The primary therapy period of this study was from 25-May-2010 to 22-Nov-2011.

The original protocol (amendment 2) was dated 18 February 2010 and there were protocol clarification letters dated 05 May 2010, 31 March 2010 and 01 Aug 2011. Changes to the statistical analysis plans as detailed in the protocol were made in memos dated 31 August 2011 and 8 December 2011.

##### 3.1.1.1 Study Design and Statistical Methods

###### Objectives:

###### Maintenance

1. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by change from baseline in mean subjective total sleep time (sTSTm) on the daily sleep e-diary at Month 1.
2. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by the change from baseline in wakefulness after persistent sleep onset (WASO) at Month 1.
3. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by change from baseline in mean subjective total sleep time (sTSTm) on the daily sleep e-diary at Month 3.
4. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by the change from baseline in wakefulness after persistent sleep onset (WASO) at Month 3.

###### Onset

5. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by change from baseline in mean subjective time to sleep onset (sTSOm) on the daily sleep e-diary at Month 1.

6. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by the change from baseline in latency to onset of persistent sleep (LPS) at Month 1.
7. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by change from baseline in mean subjective time to sleep onset (sTSOm) on the daily sleep e-diary at Month 3.
8. To evaluate the efficacy of MK-4305 high dose compared with placebo in improving insomnia, as measured by the change from baseline in latency to onset of persistent sleep (LPS) at Month 3.

This was a randomized, double-blind (with in-house blinding), placebo-controlled, parallel group, multicenter questionnaire and PSG study to assess the safety, tolerability and efficacy of MK-4305 in the treatment of patients with Primary Insomnia. The overall study was comprised of a screening period (including a 2-week single-blind placebo run-in), a 3-month double-blind core treatment period, an optional 3-month double-blind extension, and a 1-week double-blind run-out period at the conclusion of a patient's treatment (occurring either after the core treatment period or after completion of the extension period for participating patients). Patients were recruited to either the Questionnaire-only cohort (Q-cohort) or the PSG-plus-Questionnaire cohort (PQ-cohort). The Sponsor was to identify which sites were to be enrolling patients into the individual cohorts. Patients in both cohorts were to complete a daily sleep questionnaire via an electronic diary (e-diary). Patients in the PQ-cohort were to additionally undergo PSG assessments.

The run-in period was to commence at Visit 2 and continue until the patient returned at Visit 3. Patients who continued to meet the overall and cohort-specific inclusion/exclusion criteria at Visit 3 were to be randomized in a 3:2:3 ratio to receive either MK-4305 [high dose], MK-4305 [low dose], or placebo. The dose of MK-4305 received was to be determined by age group as follows:

Non-elderly (18 to < 65 years): MK-4305 high and low doses were to be 40 and 20 mg, respectively.

Elderly (65 years): MK-4305 high and low doses were to be 30 and 15 mg, respectively.

During the core treatment period, all patients were to return to the clinic after randomization, for visits at the end of Week 2, and the end of Months 1, 2, and 3. For patients in the PQ cohort, the visits at Night 1 and end of Months 1 and 3 were to be overnight PSG visits. Q-cohort was to advance to Visit 2 and Visit 3 to assess for continued patient eligibility prior to randomization. PQ-cohort was to advance to Visit 2-PSG: Screening PSG and Visit 2a-PSG: Baseline PSG, during which specific PSG criteria must be met and exclusionary sleep disorders ruled out at each respective overnight visit.

### **Efficacy Analyses**

The primary hypotheses were to be evaluated by comparing the MK-4305 high dose to placebo for maintenance endpoints: change from baseline in mean subjective total sleep time (sTSTm) and change from baseline in wakefulness after persistent sleep onset (WASO) at Months 1 and 3; and for onset endpoints: change from baseline in mean subjective time to sleep onset (sTSOm) and change from baseline in latency to onset of persistent sleep (LPS) at Months 1 and 3, while

the high dose secondary hypotheses were to be evaluated for the same endpoints at Week 1 (sTSTm and sTSOm) or Night 1 (WASO and LPS). Statistical significance for the primary hypotheses was to be based on the following multiplicity strategy to control the overall Type I error at the two-sided 5% significance level: to account for the evaluation of two distinct indications a Bonferroni approach was to be used; within each indication, a fixed sequential testing procedure was to be used to move from the first set of primary hypotheses (Month 1) to the next set of primary hypotheses (Month 3). Within each time point, a Hochberg approach was to be used to evaluate the objective and subjective endpoints.

### **Power and Sample Size**

This study was to randomize 360 patients into the MK-4305 high dose and placebo groups, and 240 patients into the MK-4305 low dose group (total N of 960); these are the sample sizes that were to be used to evaluate the questionnaire endpoints. This includes patients in the PQ-cohort (75% of sample size) as well as the Q-cohort (25% of sample size). For the PSG endpoints (i.e., collected in the PQ-cohort), a total of 270 patients were to be randomized to each of the MK-4305 high dose and placebo groups, and 180 patients were to be randomized to the MK-4305 low dose. Based upon the sample sizes noted above and assuming that the overall dropout rate is approximately 1% at Night 1 (for PSG endpoints), 5% at Week 1 (for questionnaire endpoints), 10% at Month 1, and 20% at Month 3 (and similar among the treatment groups), the study would have 91% power to declare all primary maintenance endpoints significant (i.e., for MK-4305 high dose vs. placebo) in accordance with the multiplicity strategy noted above. The study also would have 62% power to declare all primary onset endpoints significant (i.e., for MK-4305 high dose vs. placebo); the probability of declaring both Month 1 onset endpoints significant and at least one Month 3 onset endpoint significant is 81%. These probabilities are based upon estimates of standardized effect sizes (reduced by 10% to account for study variability) from Period 1 of the MK- 4305 Phase IIb study (Protocol 006).

### **Efficacy Analysis Populations**

The PSG Full Analysis Set (FAS-PSG) population was to serve as the primary population for the analysis of PSG efficacy data. The FAS-PSG population consists of all randomized patients who have: at least one post-randomization PSG observation subsequent to administration of at least one dose of study treatment; baseline data for those analyses that require baseline data.

The e-diary Full Analysis Set (FAS–e-diary) population was to serve as the primary population for the analysis of e-diary efficacy data. The FAS–e-diary population consists of all randomized patients who have: at least one post-randomization e-diary observation (i.e., weekly mean) subsequent to administration of at least one dose of study treatment; baseline data for those analyses that require baseline data. Note that the number of patients included in the FAS populations may vary across endpoints due to the degree of missing data for each endpoint.

### **Changes in Planned Analyses**

Changes to the analysis were prespecified in two protocol clarification letters (dated 8/31/2011 and 12/8/2011).

Based upon the sponsor's blinded data review there were some questionable e-diary daily patient data where sTSO+sTST+sWASO was more than 24 hours which the sponsor thought may lack reliability. These records were therefore to be removed prior to deriving the weekly means for each variable in the e-diary. The number of such occurrences was to be summarized separately.

Handling of patients enrolling multiple times (within, or between P028 and P029) was to be as follows.

Sequential Enrollments for Same Subject: The first subject ID was to be included in any summaries or analyses of data involving the FAS dataset. Subsequent subject IDs (for this same patient) were to be excluded from all of the analysis datasets since these data do not represent an independent assessment of efficacy or safety.

Overlapping Enrollments for Same Subject Within Same Study or In Different Studies: All subject IDs from this patient were to be excluded from all of the analysis datasets since there are potential questions regarding the validity (e.g., patient fraud) of data from such a patient.

Definition of Sleep Onset Latency (SOL) in the protocol was changed to: 'the duration of time measured from lights off to the first epoch of 3 consecutive stage S1 or any epoch of stage S2, S3, or stage R,' to be in accordance with current practice of the central scorer.

For the primary efficacy analysis model, region is five levels for subjective endpoints and 4 levels for objective endpoints. Prior to unblinding, it was decided to define region according to the protocol rather than the follow-up Protocol Clarification Letter.

The day ranges for the three post-baseline PSG assessments are defined as follows: Night 1 (Day 1), Month 1 (+/-10 days from the target Day 30), and Month 3 (+/-14 days from the target Day 89). The protocol had specified +/-7 days from the target dates at Months 1 and 3; however, during blinded data review it was noted that multiple observations would have been excluded if these tighter day ranges were used, so a decision was made prior to unblinding to use the wider ranges noted. Primary and secondary hypotheses related to PSG endpoints were also conducted using the protocol-specified day ranges to evaluate robustness of results.

## **Derivations of Efficacy Endpoints**

### **PSG Data**

Sleep stage scoring of the PSG recordings will be performed according to Rechtschaffen and Kales (R&K) criteria for each 30-second epoch by a central sleep scoring laboratory. Each 30-second epoch will be scored as wake, Stage 1, 2, 3, 4 or REM. The primary, secondary and exploratory PSG endpoints for analysis will be derived from this information.

### **Questionnaire Data**

For the morning and evening questionnaire data, "weekly" averages will be calculated. Questionnaire data provided on the nights in the sleep lab will be excluded from the derivation of these weekly averages since: a) the subjective endpoints may be influenced by being at the sleep lab versus in the outpatient environment; b) the time in bed is limited to 8 hours in the sleep lab. Specifically, weekly averages will be derived by taking the mean of all available daily measurements (excluding the mornings following any PSG nights) falling within the

corresponding day range noted below. Note that patients must have at least 3 days of data during each week to calculate a weekly average; otherwise, the mean value will be considered missing for that week/month. Also note that if a patient ends the extension period earlier than the target day (180), then up to last 14 days (after day 155) leading to the end of the extension period will be the day range for Month 6. The baseline value will be the mean of the last 7 (non-missing) measurements obtained during the placebo run-in period.

### **Patients Who Do Not Fall Asleep and Other Conventions**

For the PSG, total sleep time (TST) may be recorded as 0; the frequency of such cases will be summarized in the report. In this case, the patient likely did not fall asleep for a whole night. Since the at-risk time for WASO and NAW is 0, WASO and NAW are undefined and will be set as missing. The best result for LPS is 0 minutes. When a patient does not fall asleep, it is the worst outcome. Since the at-risk time for LPS is the duration of time in bed (TIB), which is 8 hours for PSG nights, the LPS will be set as 8 hours if TST=0.

For the e-diary, if the answer to the question "Did you fall asleep at all last night?" is "no" (i.e., sTST=0) and is confirmed by a follow-up question, then imputations for morning e-diary endpoints were to be employed as noted below for the same reasons described above. Additionally, a "worst value" of 1 was to be imputed for sQUAL when sTST=0. sTSO = 480 minutes sNAW = missing sWASO = missing sQUAL = 1 (i.e., "Poor") sNAW may be recorded as 0. When sTST>0, sNAW=0, and sWASO is missing, sWASO was to be imputed as 0.

### **Statistical Methods for Efficacy Analyses**

For the analysis of change from baseline in sTST<sub>m</sub>, WASO, sTSO<sub>m</sub>, and LPS, a longitudinal data analysis(LDA) method was to be used. This model assumes a different mean for each treatment at each of the repeated time points in the analysis. In this model, time is treated as a categorical variable so that no restriction is imposed on the trajectory of the means over time. The analysis model was to adjust for baseline value of the response variable (if applicable), age group (non-elderly vs. elderly), region, gender, treatment, time, and the interaction of treatment by time.

To evaluate efficacy hypotheses (e.g., LPS at Month 1), efficacy data were to be included in the model for all time points assessed during the 3-month treatment period; however, appropriate contrasts were to be used to test the treatment difference of interest (e.g., at Month 1). An unstructured covariance matrix was to be used to model the correlation among repeated measurements.

To evaluate the robustness of the efficacy findings based on LDA, a nonparametric approach using Multiple Imputation followed by an aligned rank analysis (Hodges and Lehmann and Mogg and Mehrotra) was to be performed as a sensitivity analysis for the primary and secondary endpoints. If inconsistent results were observed between the primary analysis (LDA) and the sensitivity analysis (the nonparametric approach mentioned above) for a particular primary or secondary endpoint, an additional nonparametric approach, the ETRANK procedure, was to be used as a sensitivity analysis to analyze the incomplete repeated measures data for that endpoint to evaluate the effect of drop-outs on the treatment difference. The ETRANK method uses the observed data (without imputation or estimating the missing data) or the endpoint data and

creates scoring systems that are either categorical, time-related ranks or the observed levels. It was designed for treatment-related patterns of withdrawal.

For the analysis of responders (e.g., based on change from baseline in ISI total score 6 points), a generalized linear mixed model was to be used. This model assumes a binary distribution for the response and uses a logit link. The treatment difference in terms of log odds ratio was to be estimated and tested from this model using the SAS PROC GLIMMIX procedure. All other aspects of the responder analysis are analogous to the primary analysis for ISI total score.

### **DETAILS OF NON-PARAMETRIC SENSITIVITY ANALYSIS**

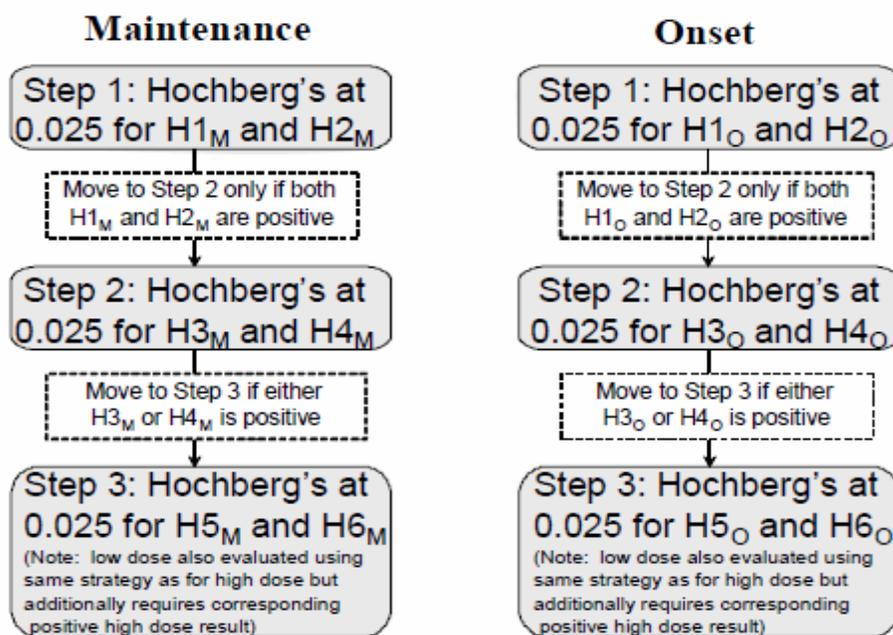
**Step 1 (handling missing data):** Ten multiple imputed complete data sets were to be constructed using regression where a regression model is fitted to each variable with missing values with previous variables as covariates. In other words, imputation was to be carried out sequentially over time (e.g., Night 1, Month 1, then Month 3 for PSG endpoints) for the time points at which imputation is required separately for each treatment group. SAS PROC MI was to be used to implement the imputation procedure which assumes the missing data are missing at random (MAR) and limits imputation to monotone missing patterns. If there was a non-monotone missing pattern then the intermittent missing data (expected to be a relatively small percentage of the missing data) was to be imputed by carrying the last observed data forward prior to the multiple imputation. For each multiple imputed aligned data set, Step 2 through Step 4 were to then be performed. **Step 2 (alignment to adjust for covariates):** Patients were to be grouped according to the following covariates: baseline (dichotomous at the median), age (non-elderly, elderly), region, and gender (and Q-cohort/PQ-cohort for diary endpoints). Each intersection of these covariates was to be considered as a block. At each particular time point, the data was to be centered by subtracting the block median from each observation within each block. The centered data are now "aligned" or adjusted for all covariates (Hodges and Lehmann). **Step 3 (assigning ranks):** at each time point ranks were to be assigned to the aligned data. **Step 4 (rank analysis):** Wilcoxon sum rank test was to be performed on the ranks. **Step 5 (combine results):** The 10 Wilcoxon sum rank tests from the 10 multiple imputed aligned data sets were to be combined using SAS PROC MIANALYZE.

### **Multiplicity**

While nominal p-values were to be computed for all comparisons of MK-4305 high dose with placebo, statistical significance for the primary and secondary hypotheses was to be based on the following multiplicity strategy. To account for the evaluation of two distinct indications a Bonferroni approach was to be used; that is, endpoints evaluated to assess the sleep maintenance effect (sTSTm and WASO) were to be tested at the two-sided 2.5% level and endpoints evaluated to assess the sleep onset effect (sTSOm and LPS) were to be tested at the two-sided 2.5% level. Within each indication, a fixed sequential testing procedure was to be used to move from the first set of primary hypotheses (Month 1) to the next set of primary hypotheses (Month 3) Within each time point, a Hochberg approach was to be used to evaluate the subjective (e.g., sTSTm) and objective (e.g., WASO) endpoints; however, to move sequentially from Month 1 to Month 3 (as noted above), both the subjective and objective endpoints needed to be significant according to this procedure. If only one of the endpoints at Month 1 was significant, then that endpoint was declared positive, but the testing procedure for the indication stopped and no further conclusions could be made regarding the effect of MK-4305 high dose at Month 3. Hence the overall Type I error among all primary hypotheses was to be controlled at the two-sided 5%

significance level. In addition, statistical significance for the high dose secondary hypotheses within each indication was to be based on the following: if either Month 3 hypothesis was positive (sTST<sub>M</sub> or WASO for maintenance indication and sTSO<sub>M</sub> or LPS for onset indication) according to the multiplicity strategy described above, then the set of secondary hypotheses (Week 1/Night 1) was to be tested using a Hochberg approach at the two-sided 2.5% level. This strategy is further illustrated in Figure 1 below. Low dose comparisons to placebo within each indication for the primary and secondary endpoints were to be evaluated if at least one of the Month 3 endpoints was positive for the high dose (according to the multiplicity strategy noted above). Evaluation of these low dose hypotheses was to follow the same multiplicity strategy as noted for the high dose with one additional requirement: for a particular endpoint, the high dose must be positive in order to declare the low dose positive.

**Figure 1 Multiplicity Adjustment Method used in Phase 3 Studies**



Notes: This figure was copied from page 113 of the protocol

H1<sub>M</sub>: High Dose – Placebo=0 for sTST at Month 1 ; H2<sub>M</sub>: High Dose – Placebo=0 for WASO at Month 1

H3<sub>M</sub>: High Dose – Placebo=0 for sTST at Month 3 ; H4<sub>M</sub>: High Dose – Placebo=0 for WASO at Month 3

H5<sub>M</sub>: High Dose – Placebo=0 for sTST at Night 1 ; H6<sub>M</sub>: High Dose – Placebo=0 for WASO at Night 1

To get the corresponding Onset hypotheses replace sTST with sTSO and WASO with LPS in the Maintenance ones.

While the lower dose of MK-4305 was also of interest, the protocol stated that it was exploratory in this study due to the reduced allocation; the lower dose was to be compared to placebo using a similar multiplicity strategy in a pooled analysis including this study and another similarly designed Phase III efficacy study.

**Reviewer's Comment:**

Contrary to the sponsor's use of the conjunction 'or' in the boxes under step 2 in Figure 1, in order to strongly control the type I error at the noted level both hypotheses in each step would need to be significant.

**3.1.1.2 Patient Disposition**

Of the 1022 patients randomized into the study, 254 and 383 patients were randomized to suvorexant LD and suvorexant HD, respectively, and 385 were randomized to placebo. Of these, one patient (AN 07264) was randomized and did not take any assigned study therapy (placebo group); therefore, 1021 patients received at least one dose of study drug and were evaluated for safety. This patient discontinued due to physician decision. A total of 916 (89.6%) patients completed the Treatment Phase and 105 (10.3%) discontinued the study. Note that three patients attended the End of Month 3 visit but did not complete all assessments.

The proportion of patients who discontinued during the Treatment Phase was similar between placebo and the suvorexant treatment groups and ranged from 9.4% to 11.2%. The most common reason for discontinuation in the Treatment Phase (as well as in both age groups) was due to an AE, with incidences ranging from 2.4% to 3.9% for suvorexant and 5.5% for placebo. The discontinuation rate among treatment groups was similar for non-elderly and elderly patients.

**Table 2 Study 28: Patient Disposition**

	MK-4305 LD n (%)	MK-4305 HD n (%)	Placebo n (%)	Total n (%)
<b>Not Randomized</b>				1856
<b>Patients in population</b>	254	383	385	1022
<b>Study Disposition</b>				
Not Treated	0 (0.0)	0 (0.0)	1 (0.3)	1 (0.1)
Completed Treatment <sup>†</sup>	230 (90.6)	345 (90.1)	341 (88.6)	916 (89.6)
Discontinued during Treatment	24 (9.4)	38 (9.9)	43 (11.2)	105 (10.3)
Adverse Event	6 (2.4)	15 (3.9)	21 (5.5)	42 (4.1)
Withdrawal by Subject	6 (2.4)	8 (2.1)	12 (3.1)	26 (2.5)
Protocol Violation	5 (2.0)	3 (0.8)	1 (0.3)	9 (0.9)
Lost to Follow-up	1 (0.4)	1 (0.3)	0 (0.0)	2 (0.2)
Lack of Efficacy	1 (0.4)	7 (1.8)	9 (2.3)	17 (1.7)
Pregnancy	1 (0.4)	1 (0.3)	0 (0.0)	2 (0.2)
Physician Decision	4 (1.6)	3 (0.8)	0 (0.0)	7 (0.7)
<b>Protocol Milestone</b>				
Completed Treatment	230 (90.6)	345 (90.1)	341 (88.6)	916 (89.6)
Continuing Into Extension	100 (39.4)	172 (44.9)	151 (39.2)	423 (41.4)
Continuing Into Run-Out	128 (50.4)	172 (44.9)	186 (48.3)	486 (47.6)
Not Treated in Run-Out	0 (0.0)	1 (0.3)	0 (0.0)	1 (0.1)
Not Continuing Into Extension Or Run-Out <sup>†</sup>	2 (0.8)	1 (0.3)	4 (1.0)	7 (0.7)
Discontinued during Treatment (not cont into Ext or RO)	24 (9.4)	38 (9.9)	43 (11.2)	105 (10.3)
Each patient is counted once for Study Disposition and Protocol Milestone based on the latest corresponding disposition record.				
<sup>†</sup> AN 7156, 7073, and 8052 attended V7 but did not complete all assessments.				
MK-4305 LD = MK-4305 20 mg for patients < 65 years and MK-4305 15 mg for patients ≥ 65 years.				
MK-4305 HD = MK-4305 40 mg for patients < 65 years and MK-4305 30 mg for patients ≥ 65 years.				

Note: This table was copied from the sponsor's study report, page 201

BEST AVAILABLE COPY

### 3.1.1.3 Baseline Demographics and Disease Characteristics

Approximately two-thirds of patients were female; further, the gender distribution across treatment groups was generally similar. The ages ranged from 18 to 87 years, with a mean age of 56 years. Of the treated patients, 429 (42.0%) were elderly. The distribution of age was similar among the treatments groups. Patients were predominantly White, non-Hispanic or Latino and enrolled from sites in North America or Europe. By design more patients were enrolled into the PSG and Questionnaire (PQ) cohort (76.0%); the proportion of patients in each treatment group was similar within each cohort. Baseline characteristics for the non-elderly and elderly patients were generally similar. Baseline characteristics were generally comparable across the cohorts with the exception of race (as the Questionnaire only [Q] cohort was enrolled exclusively by sites in Japan) and body mass index (BMI). The distribution of BMI for the Q-cohort was skewed toward normal and underweight (BMI < 25: 78.9%) as compared to the PQ-cohort (BMI < 25: 42.4%); this observation is likely due to the fact that the Q-cohort was comprised exclusively of patients from Japan. The treatment groups were generally comparable with regard to baseline values of efficacy measures.

Table 3 summarizes baseline demographics of randomized patients and Table 4 summarizes efficacy measures at baseline.

**Table 3 Study 28: Baseline Demographics (Reviewer's Analysis)**

Variable	Levels	Placebo	LD	HD	All
age	Mean (SD)	56.1 (15.0)	55.2 (15.7)	55.7 (15.3)	55.7 (15.3)
agegrp		.	.	.	. (. )
agegrp	<65	224 (58.2)	148 (58.0)	222 (58.0)	594. (58.1)
agegrp	≥65	161 (41.8)	107 (42.0)	161 (42.0)	429 (41.9) .
bmi	Mean (SD)	25.2 (4.2)	25.2 (4.2)	25.1 (4.1)	25.1 (4.1)
cohort	Q	94 (24.4)	61 (23.9)	92 (24.0)	247 (24.1)
cohort	PQ	291 (75.6)	194 (76.1)	291 (76.0)	776 (75.9)
ethnicity	Hispanic or Latino	37 (9.6)	36 (14.1)	42 (11.0)	115 (11.2)
ethnicity	Not Hispanic or Latino	348 (90.4)	219 (85.9)	341 (89.0)	908 (88.8)
sex	F	246 (63.9)	163 (63.9)	230 (60.1)	639 (62.5)
sex	M	139 (36.1)	92 (36.1)	153 (39.9)	384 (37.5)
race	ASIAN	99 (25.7)	66 (25.9)	98 (25.6)	263 (25.7)
race	BLACK OR AFRICAN AMERICAN	25 (6.5)	15 (5.9)	18 (4.7)	58 (5.7)
race	MULTI- RACIAL	15 (3.9)	5 (2.0)	14 (3.7)	34 (3.3)
race	NATIVE HAWAIIAN OR OTHER PACIF	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
race	WHITE	245 (63.6)	169 (66.3)	253 (66.1)	667 (65.2)
Region	Asia/ Eastern Europe/ Africa	18 (4.7)	9 (3.5)	12 (3.1)	39 (3.8)
region	Europe	134 (34.8)	89 (34.9)	135 (35.2)	358 (35.0)
region	Japan	94 (24.4)	61 (23.9)	92 (24.0)	247 (24.1)
region	North America	125 (32.5)	92 (36.1)	129 (33.7)	346 (33.8)
region	Other Cent South America	14 (3.6)	4 (1.6)	15 (3.9)	33 (3.2)

**Table 4 Study 28: Baseline Efficacy Measures (Reviewer’s Analysis)**

Variable	Levels	Placebo	LD	HD	All
LPS	Mean (SD)	66.2 (44.1)	68.9 (49.7)	61.8 (39.1)	65.2 (43.8)
TST	Mean (SD)	307.5 (63.5)	299.8 (62.8)	308.3 (65.5)	305.9 (64.1)
WASO	Mean (SD)	114.9 (45.7)	119.2 (46.5)	117.7 (49.6)	117.0 (47.4)
sTSoM	Mean (SD)	66.9 (40.5)	63.3 (37.1)	68.0 (50.1)	66.4 (43.6)
sTSTm	Mean (SD)	315.7 (65.1)	322.6 (57.3)	316.1 (67.2)	317.6 (64.1)
sWASOPm	Mean (SD)	119.5 (77.0)	117.6 (68.6)	118.7 (68.5)	118.7 (71.8)
sWASOm	Mean (SD)	78.2 (52.5)	73.8 (45.0)	78.4 (50.7)	77.2 (50.1)

### 3.1.1.4 Sponsor’s Results

Of 1023 patients randomized to the three treatment groups, one patient (AN 07126 noted below) enrolled in more than one suvorexant study in an overlapping fashion and these data were subsequently excluded from all summaries and analyses according to rules documented prior to unblinding. Excluding this one patient, 1022 patients were randomized in this study; this population constitutes the All Patients Randomized (APR) set.

#### Subjective Sleep Maintenance –sTSTm

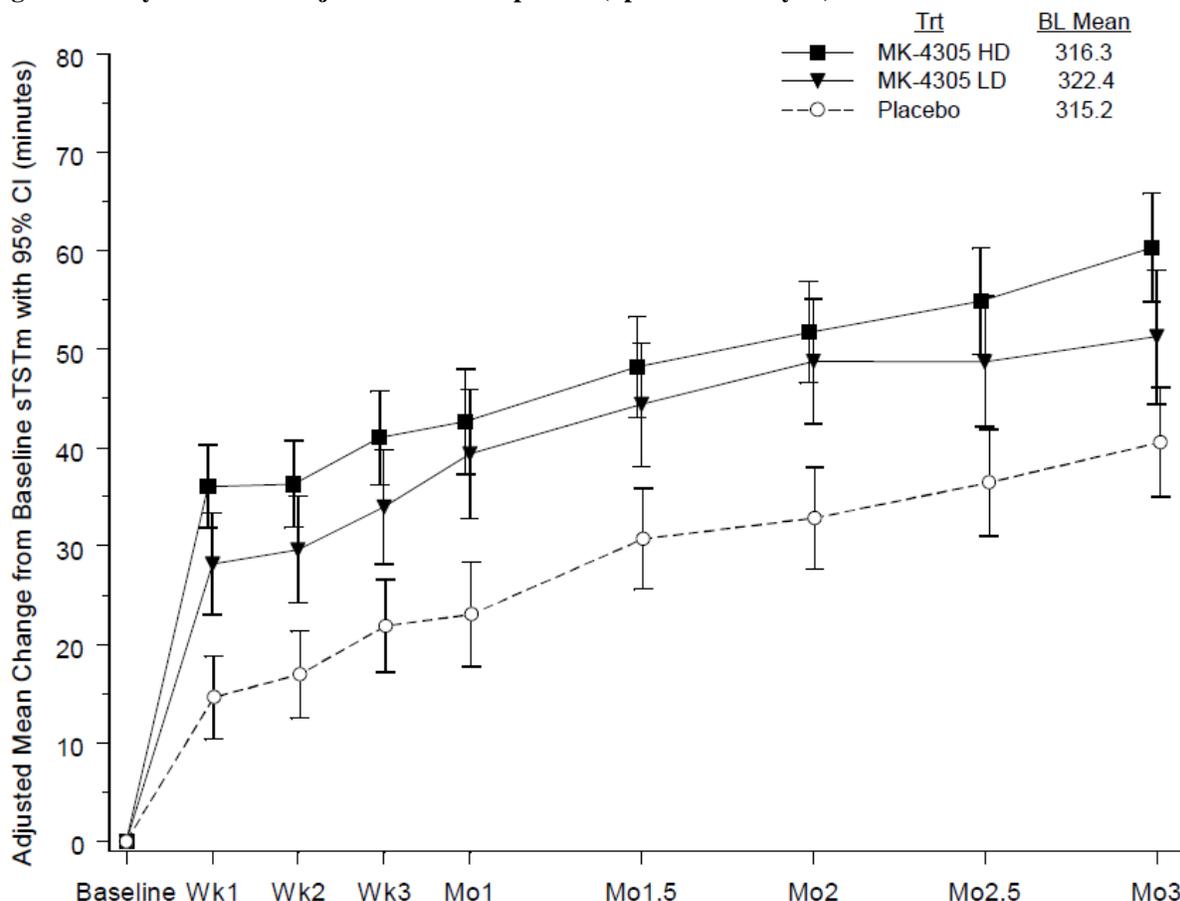
The analyses shown in Table 5 and Figure 2 suggest that suvorexant HD was superior to placebo in increasing subjective total sleep time at the primary (Month 1 and Month 3) timepoints and also at the secondary (Week 1) timepoint (all p-values < 0.00001). The improvements in sTSTm from baseline for suvorexant HD ranged from 36.0 to 60.3 minutes on average. Patients on suvorexant HD improved on average 19.6 to 21.4 minutes more than placebo patients. Nominal (and multiplicity-adjusted) p-values suggest that suvorexant LD was more efficacious than placebo in increasing sTSTm at the primary and secondary timepoints (all p-values < 0.025). The improvements in sTSTm from baseline for suvorexant LD ranged from 28.2 to 51.2 minutes on average. Patients on suvorexant LD improved 10.7 to 16.3 minutes more than placebo.

**Table 5 Study 28: Mean Subjective Total Sleep Time (Sponsor’s Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) †
<b>Week 1</b>					
MK-4305 LD	248	322.4 ( 57.3)	349.6 ( 59.0)	27.2 ( 40.8)	28.2 ( 23.0, 33.4)
MK-4305 HD	379	316.3 ( 67.4)	352.6 ( 72.0)	36.3 ( 46.0)	36.0 ( 31.8, 40.2)
Placebo	376	315.2 ( 65.2)	330.4 ( 67.5)	15.3 ( 42.9)	14.6 ( 10.4, 18.8)
<b>Month 1</b>					
MK-4305 LD	244	322.7 ( 57.7)	361.5 ( 66.3)	38.7 ( 50.5)	39.4 ( 32.8, 45.9)
MK-4305 HD	363	317.6 ( 64.0)	361.9 ( 72.3)	44.2 ( 57.8)	42.6 ( 37.3, 48.0)
Placebo	365	317.7 ( 65.3)	341.0 ( 69.6)	23.4 ( 52.0)	23.1 ( 17.7, 28.4)
<b>Month 3</b>					
MK-4305 LD	228	325.4 ( 56.7)	375.7 ( 61.3)	50.3 ( 55.2)	51.2 ( 44.4, 58.1)
MK-4305 HD	348	316.6 ( 65.8)	378.9 ( 72.4)	62.2 ( 58.0)	60.3 ( 54.8, 65.8)
Placebo	339	316.7 ( 64.5)	358.8 ( 64.8)	42.1 ( 56.4)	40.6 ( 35.0, 46.1)
Pairwise Comparison		Difference in LS Means (95% CI)†			p-Value†
<b>Week 1</b>					
MK-4305 HD vs. Placebo		21.4 ( 15.5, 27.4)			<0.00001
MK-4305 LD vs. Placebo		13.6 ( 6.9, 20.3)			0.00007
<b>Month 1</b>					
MK-4305 HD vs. Placebo		19.6 ( 12.0, 27.1)			<0.00001
MK-4305 LD vs. Placebo		16.3 ( 7.9, 24.8)			0.00016
<b>Month 3</b>					
MK-4305 HD vs. Placebo		19.7 ( 11.9, 27.6)			<0.00001
MK-4305 LD vs. Placebo		10.7 ( 1.9, 19.5)			0.01711
† Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates.					
MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years.					
MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 259 of the sponsor’s study report

Figure 2 Study 28: Mean Subjective Total Sleep Time (Sponsor's Analysis)



Note: This figure was copied from page 260 of the sponsor's study report

### Subjective Sleep Onset – sTSOM

Table 6 presents summary statistics and analysis results for sTSOM at the key timepoints during the Treatment Phase for both the HD and LD suvorexant. The analyses show that suvorexant HD was superior to placebo in decreasing sTSOM at the primary (Month 1 and Month 3) timepoints and also at the secondary (Week 1) timepoint (all p-values < 0.01). The improvements in sTSOM from baseline for suvorexant HD ranged from 15.3 to 25.7 minutes on average. Patients on suvorexant HD improved 5.7 to 8.4 minutes more than those on placebo. Nominal (but not multiplicity-adjusted) p-values provide evidence to suggest that suvorexant LD was more efficacious than placebo in decreasing sTSOM at Month 3 and Week 1. The improvements from baseline in sTSOM for suvorexant LD ranged from 15.2 to 22.5 minutes on average. Patients on suvorexant LD improved 5.2 to 5.6 minutes more than placebo.

**Table 6 Study 28: Mean subjective Time to Sleep Onset (Sponsor's Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) †
<b>Week 1</b>					
MK-4305 LD	248	63.6 ( 37.3)	49.1 ( 33.1)	-14.5 ( 28.4)	-15.2 (-18.7, -11.7)
MK-4305 HD	379	67.9 ( 50.2)	51.7 ( 44.7)	-16.2 ( 32.7)	-15.3 (-18.1, -12.4)
Placebo	376	67.2 ( 40.7)	57.1 ( 40.1)	-10.1 ( 33.9)	-9.6 (-12.5, -6.7)
<b>Month 1</b>					
MK-4305 LD	244	62.7 ( 36.7)	46.3 ( 32.0)	-16.4 ( 31.5)	-17.1 (-21.4, -12.9)
MK-4305 HD	363	65.3 ( 41.2)	44.6 ( 32.1)	-20.7 ( 36.9)	-19.1 (-22.6, -15.7)
Placebo	365	65.7 ( 39.4)	52.9 ( 40.7)	-12.8 ( 41.2)	-11.7 (-15.2, -8.2)
<b>Month 3</b>					
MK-4305 LD	228	60.5 ( 34.7)	40.1 ( 28.8)	-20.4 ( 27.5)	-22.5 (-26.3, -18.7)
MK-4305 HD	348	66.4 ( 45.5)	38.9 ( 33.0)	-27.4 ( 36.6)	-25.7 (-28.8, -22.6)
Placebo	339	66.6 ( 39.9)	47.7 ( 35.8)	-18.9 ( 39.3)	-17.3 (-20.4, -14.2)
Pairwise Comparison		Difference in LS Means (95% CI)†			p-Value†
<b>Week 1</b>					
MK-4305 HD vs. Placebo		-5.7 (-9.7, -1.6)			0.00609
MK-4305 LD vs. Placebo		-5.6 (-10.2, -1.1)			0.01564
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-7.4 (-12.3, -2.5)			0.00298
MK-4305 LD vs. Placebo		-5.4 (-10.9, 0.0)			0.05191
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-8.4 (-12.8, -4.0)			0.00019
MK-4305 LD vs. Placebo		-5.2 (-10.2, -0.3)			0.03771
† Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates.					
MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years.					
MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 268 of the sponsor's study report

### Objective Sleep Maintenance –WASO

Table 7 presents summary statistics and analysis results for WASO at the key timepoints during the Treatment Phase for both the HD and LD of suvorexant. The analyses show that suvorexant HD was superior to placebo in decreasing WASO at the primary (Month 1 and Month 3) timepoints and also at the secondary (Night 1) timepoint (all p-values < 0.00001). The improvements in WASO from baseline for suvorexant HD ranged from 45.0 to 58.0 minutes on average. Patients on suvorexant HD improved 22.9 to 38.4 minutes more than placebo. Nominal (and multiplicity-adjusted with sTSTm) p-values also suggest that suvorexant LD was more efficacious than placebo in decreasing WASO at the primary and secondary timepoints (all p-values < 0.0001). The improvements in WASO from baseline for suvorexant LD ranged from 41.6 to 52.1 minutes on average. Patients on suvorexant LD improved 16.6 to 32.5 minutes more than those on placebo.

**Table 7 Study 28: Mean Objective WASO (Sponsor's Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI)†
<b>Night 1</b>					
MK-4305 LD	192	119.5 (46.4)	65.3 (42.3)	-54.3 (44.7)	-52.1 (-57.4, -46.8)
MK-4305 HD	291	117.7 (49.6)	59.2 (37.3)	-58.5 (48.5)	-58.0 (-62.3, -53.7)
Placebo	287	115.1 (45.9)	96.0 (50.5)	-19.1 (47.5)	-19.6 (-23.9, -15.3)
<b>Month 1</b>					
MK-4305 LD	185	119.1 (46.0)	72.1 (45.2)	-47.0 (45.4)	-45.0 (-51.2, -38.9)
MK-4305 HD	272	116.1 (47.9)	72.0 (46.4)	-44.1 (50.1)	-45.0 (-50.1, -39.9)
Placebo	272	113.6 (45.0)	95.7 (54.5)	-17.9 (55.3)	-18.7 (-23.7, -13.6)
<b>Month 3</b>					
MK-4305 LD	172	118.2 (46.7)	75.5 (54.5)	-42.7 (50.5)	-41.6 (-48.0, -35.2)
MK-4305 HD	251	113.9 (45.3)	67.8 (42.6)	-46.1 (48.2)	-47.9 (-53.2, -42.6)
Placebo	251	115.3 (46.0)	90.0 (53.2)	-25.3 (50.7)	-25.0 (-30.3, -19.8)
Pairwise Comparison		Difference in LS Means (95% CI)†		p-Value†	
<b>Night 1</b>					
MK-4305 HD vs. Placebo		-38.4 (-44.5, -32.3)		<.00001	
MK-4305 LD vs. Placebo		-32.5 (-39.3, -25.7)		<.00001	
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-26.3 (-33.5, -19.2)		<.00001	
MK-4305 LD vs. Placebo		-26.4 (-34.3, -18.4)		<.00001	
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-22.9 (-30.3, -15.4)		<.00001	
MK-4305 LD vs. Placebo		-16.6 (-24.8, -8.3)		0.00009	
†Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates.					
MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years.					
MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from the sponsor's study report, page 265

### Objective Sleep Onset – LPS

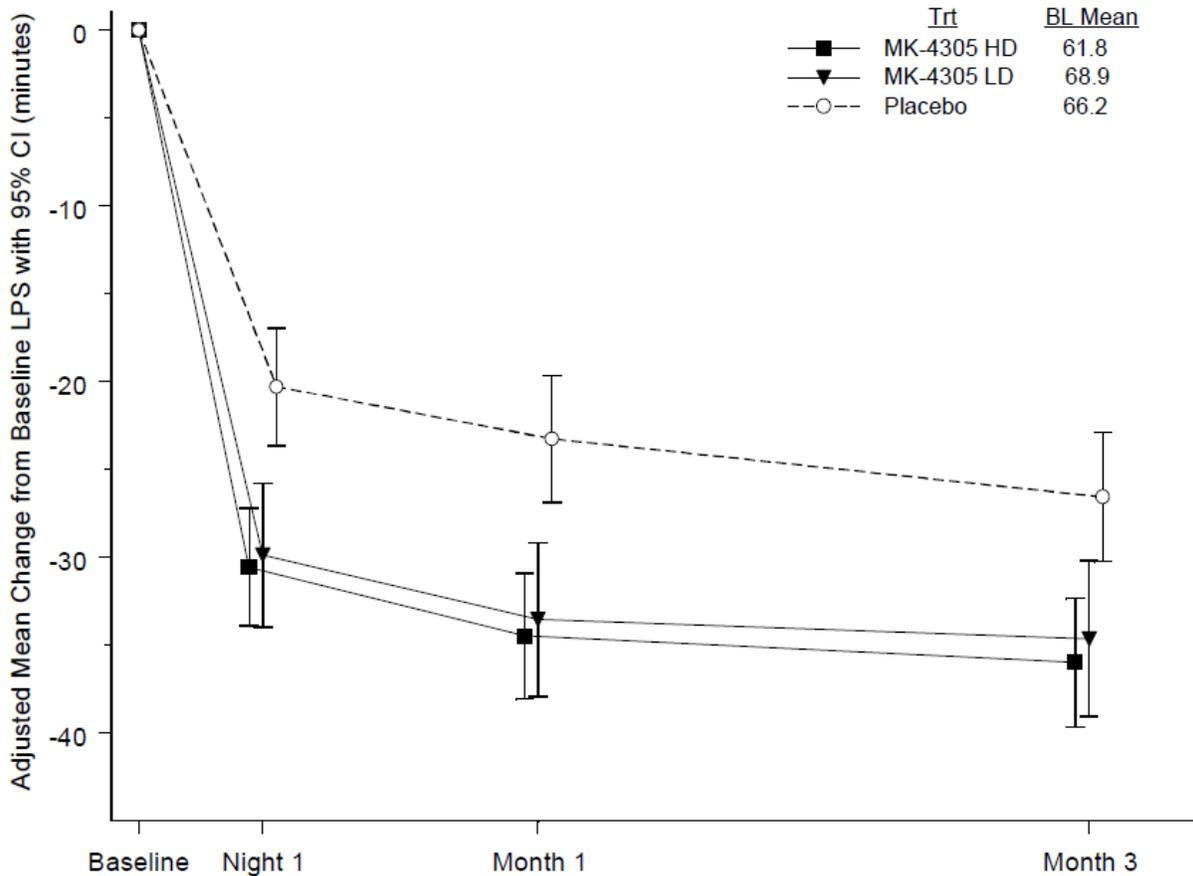
Table 8 presents summary statistics and analysis results for LPS at the key timepoints during the Treatment Phase for both the HD and LD of suvorexant(see Figure 11also). The analyses show that suvorexant HD was superior to placebo in decreasing LPS at the primary (Month 1 and Month 3) timepoints and also at the secondary (Night 1) timepoint (p-values < 0.001). The improvements in LPS from baseline for suvorexant HD ranged from 30.6 to 36.0 minutes on average. Patients on suvorexant HD improved 9.4 to 11.2 minutes more than those on placebo. Nominal (but not multiplicity-adjusted, except Month 1) p-values suggest that suvorexant LD was more efficacious than placebo in decreasing LPS at the primary and secondary timepoints. The improvements in LPS from baseline for suvorexant LD ranged from 29.9 to 34.7 minutes on average. Patients on suvorexant LD improved 8.1 to 10.3 minutes more than placebo.

Table 8 Study 28 Mean Change from Baseline in Objective LPS (Sponsor's Analysis)

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) <sup>†</sup>
<b>Night 1</b>					
MK-4305 LD	193	68.9 (49.7)	35.5 (27.8)	-33.4 (48.0)	-29.9 (-34.0, -25.8)
MK-4305 HD	291	61.8 (39.1)	33.8 (25.6)	-28.0 (41.3)	-30.6 (-33.9, -27.2)
Placebo	290	66.2 (44.1)	44.6 (37.3)	-21.6 (45.2)	-20.3 (-23.6, -17.0)
<b>Month 1</b>					
MK-4305 LD	185	67.7 (46.7)	31.7 (27.9)	-36.0 (45.5)	-33.6 (-37.9, -29.2)
MK-4305 HD	275	61.7 (39.8)	29.9 (23.7)	-31.8 (41.9)	-34.5 (-38.1, -30.9)
Placebo	272	66.2 (44.0)	41.8 (39.6)	-24.4 (51.4)	-23.3 (-26.9, -19.7)
<b>Month 3</b>					
MK-4305 LD	172	65.5 (43.7)	30.2 (24.1)	-35.2 (42.4)	-34.7 (-39.1, -30.2)
MK-4305 HD	254	61.4 (40.0)	28.0 (24.9)	-33.5 (42.7)	-36.0 (-39.7, -32.4)
Placebo	251	65.7 (43.9)	38.6 (38.5)	-27.1 (52.0)	-26.6 (-30.2, -22.9)
<b>Pairwise Comparison</b>		<b>Difference in LS Means (95% CI)<sup>†</sup></b>		<b>p-Value<sup>†</sup></b>	
<b>Night 1</b>					
MK-4305 HD vs. Placebo		-10.3 (-15.0, -5.5)		0.00002	
MK-4305 LD vs. Placebo		-9.6 (-14.9, -4.3)		0.00041	
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-11.2 (-16.3, -6.1)		0.00002	
MK-4305 LD vs. Placebo		-10.3 (-16.0, -4.6)		0.00040	
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-9.4 (-14.6, -4.3)		0.00037	
MK-4305 LD vs. Placebo		-8.1 (-13.8, -2.3)		0.00606	
<sup>†</sup> Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates. MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years. MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 271 of sponsor's study report

Figure 3 Study 28 Mean Change from Baseline in Objective LPS (Sponsor's Analysis)



Note: This figure was copied from page 272 of sponsor's study report

### Additional Analyses

The results of the aligned rank sensitivity analyses were consistent with the primary analysis for suvorexant HD at Month 1 and Week 1. However, the results of the aligned rank analyses were not statistically significant for suvorexant HD at Month 3 ( $p=0.07450$ ). As specified in the protocol, in the event that the primary and sensitivity analyses provided different conclusions, an alternative rank analysis, ETRANK, would be conducted. The ETRANK analysis is consistent with the primary analysis for suvorexant HD at Month 3. One reason the two rank analyses yield different results is that ETRANK uses an analysis that penalizes the ranks of patients who drop out due to lack of efficacy or an AE (i.e., drop-outs potentially related to treatment), and ETRANK scores that make the ranks symmetric around the median.

To further explore the comparison of suvorexant and placebo with respect to LPS, a post-hoc analysis was also performed on changes from baseline in log-transformed LPS,  $\log(1+LPS)$ . The log-transformation reduces the influence of potential outliers and departures from normality. The results of the analysis on log-transformed LPS agree with the results from the primary analysis for suvorexant HD and LD for each of the key timepoints, except suvorexant LD at Month 3.

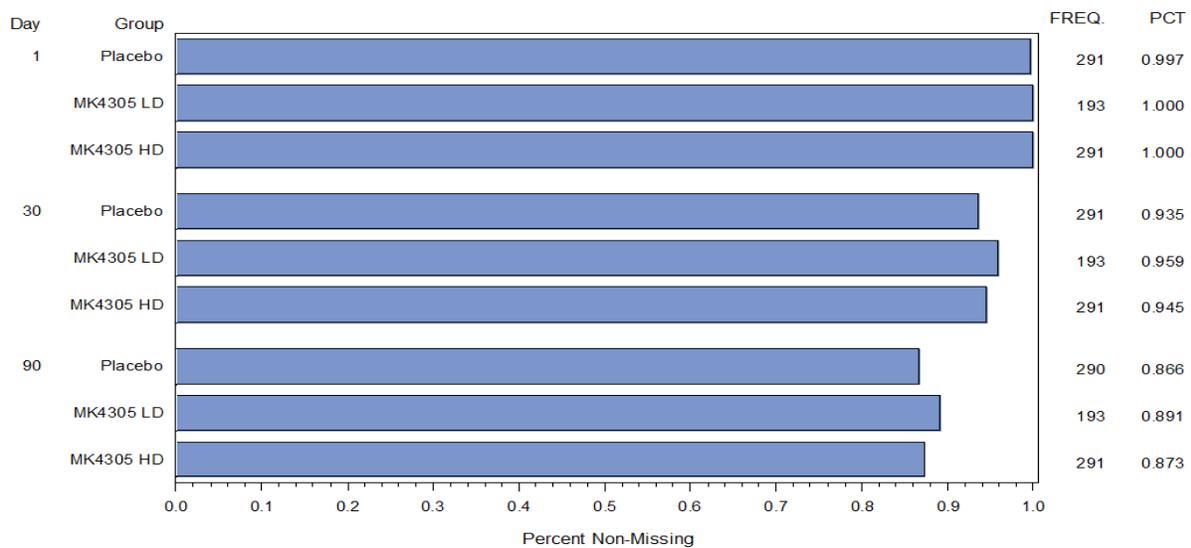
### 3.1.1.5 Reviewer's Results

This reviewer verified the sponsor's primary analyses. The reviewer did not verify the sponsor's ETRANK sensitivity analysis since it is not implemented in a SAS procedure and the sponsor didn't seem to have provided a statistical program for implementing it.

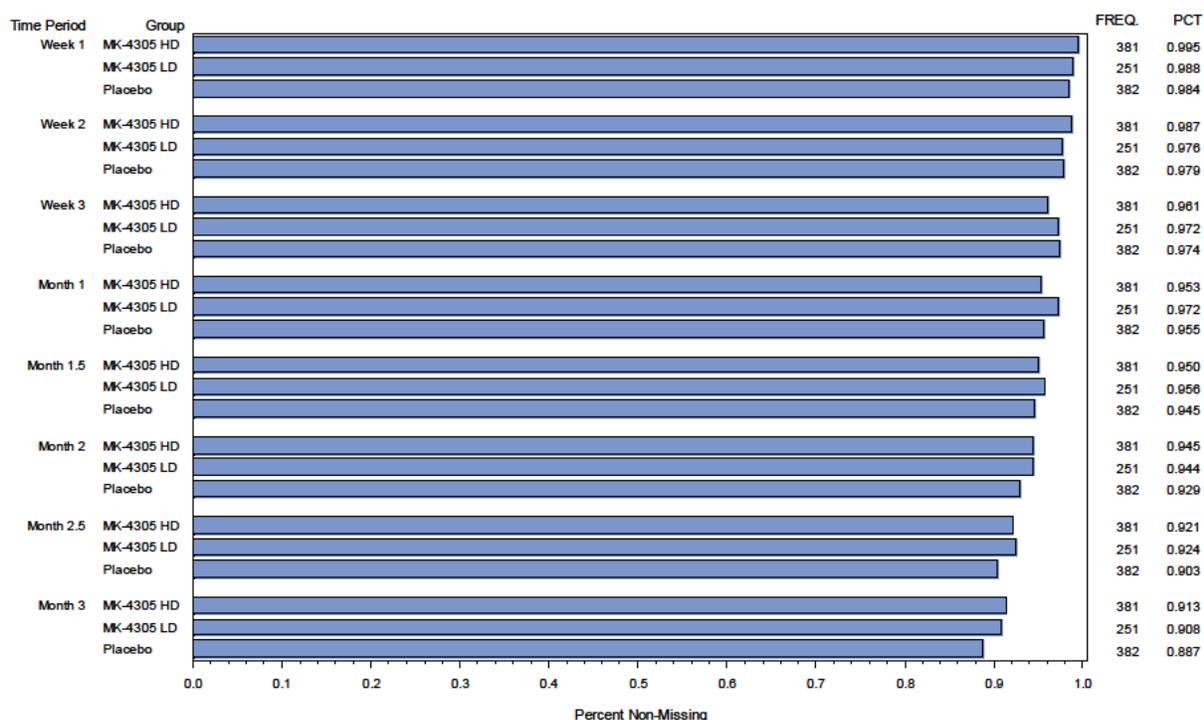
The change from baseline in LPS exhibited potentially significant non-normality. Therefore, a log transformed sensitivity analysis was undertaken. A log transformed sensitivity analysis of LPS confirmed the significance of the high dose at each timepoint (e.g., at month 1:  $p=0.0102$ ).

Figure 4 shows the extent of missing data for LPS, which was limited to about 13% at Month 3.

**Figure 4 Study 28: Percentages of Randomized Patients with non-Missing LPS data over Time**



**Figure 5 Study 28: Percentages of Randomized Patients with non-Missing sTST data over Time**



Missing data percentages were similar for the other two endpoints WASO and subjective Time to sleep onset.

Analyses of completers only were consistent with the primary analysis. For sTSO weekly averages at month 1 estimated differences from placebo for completers (N=862) were -12.9 (p<0.0001) for the high dose and -9.7 (p=0.0019) for the low dose.

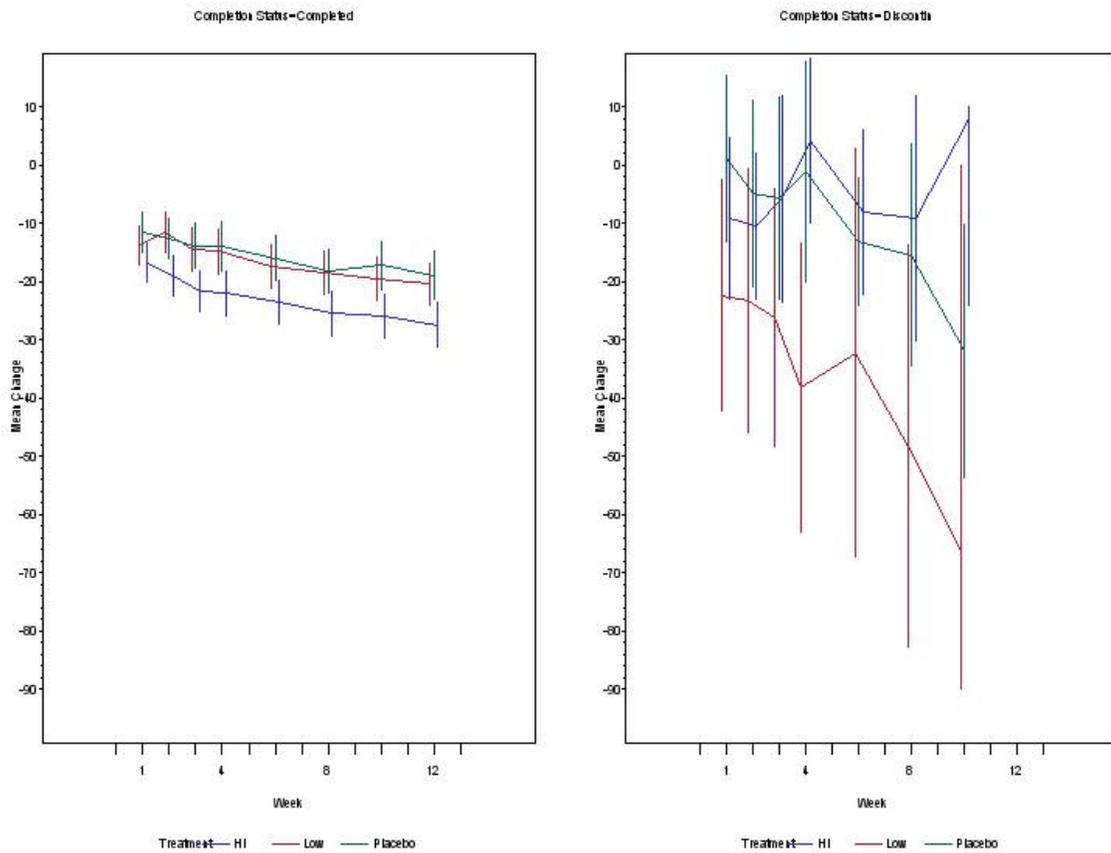
For sTSO weekly averages the low dose vs. placebo comparison (-3.46, p=0.146) was not nominally significant at month 3 based on a baseline carried forward in the event of missing data (BOCF) sensitivity analysis. Such an analysis might be a reasonable, conservative analysis for insomnia since for typical insomnia medications effects are apparent on the first night but might not be sustained later so that an LOCF analysis would be contraindicated. Analysis of completers (N=915) was not nominally significant for the low dose at month 1 (-4.2833, p=0.0729) or month 3 (-3.5084, p=0.1763). This reviewer also performed a multiple imputation sensitivity analysis for missing data. Twenty imputations were made by the Monte Carlo Markov chain method within each treatment group to get a monotone missing data pattern (no intermittent missingness). Next, for each imputation the remaining monotone missing data was imputed using a regression model developed from the placebo group data, which may be conservative when applied to the other groups. The regression model used the same covariates as the primary

analysis model as well as including any previous timepoints of the variable of interest as covariates (e.g., night 1 LPS was a covariate in the regression model for the night 30 LPS). This multiple(20) imputation analysis using an imputation model based on the placebo group was not nominally significant for the low dose vs. placebo comparison at month 1 or 3 compared to the multiplicity adjusted level of 0.025 (Month 1: -5.9,  $p=0.035$ , Month 3: -4.8  $p=0.061$ ). Figure 6 illustrates why this may be, i.e., there is a somewhat surprising phenomenon for sTSOm whereby the mean change for the low dose was very good compared to placebo (and also compared to completers) in the discontinued randomized patients subgroup. Unlike the low dose comparisons those for the high dose were still nominally significant in these sensitivity analyses. It may be worth noting that the low dose had only half the sample size of the high dose by design and all of the alpha for type I error spending was allocated to the high dose vs. placebo comparisons. At night 1 in the discontinued subgroup the low group mean LPS as measured by PSG was also markedly better than placebo but it was more comparable at the last data point available for the discontinued patient subgroup, which was month 1, thus reducing concerns about missing data bias for the analysis of LPS. An analysis of LPS for completers only ( $N=652$  for PSG) supported the results of the primary analysis (Month 1 differences: low -10.8,  $p=0.0006$ , hi: -13.0,  $p<0.0001$ ).

For sTST a completers analysis ( $N=915$ , low diff. 14.3,  $p=0.001$ ; hi diff. 20.7  $p<0.0001$ ) supported the primary analysis as did baseline carried forward for missing data or multiple imputation based on a placebo based imputation mode for all groups' missing data. The sTST pattern for discontinued subjects was similar to that for sTSO except to a lesser degree and the low group mean was not numerically better than placebo at Month 2.5 the last visit for which this subgroup had data. Therefore, there is less obvious concern for bias due to missing data for sTST.

For completeness a sensitivity analysis with patients ( $n=4$ ) reporting more than 24 hours in a day's diary data unmodified (rather than the data being excluded as prespecified) was done by this reviewer. The prespecified exclusion of this questionable data did not seem to influence the results.

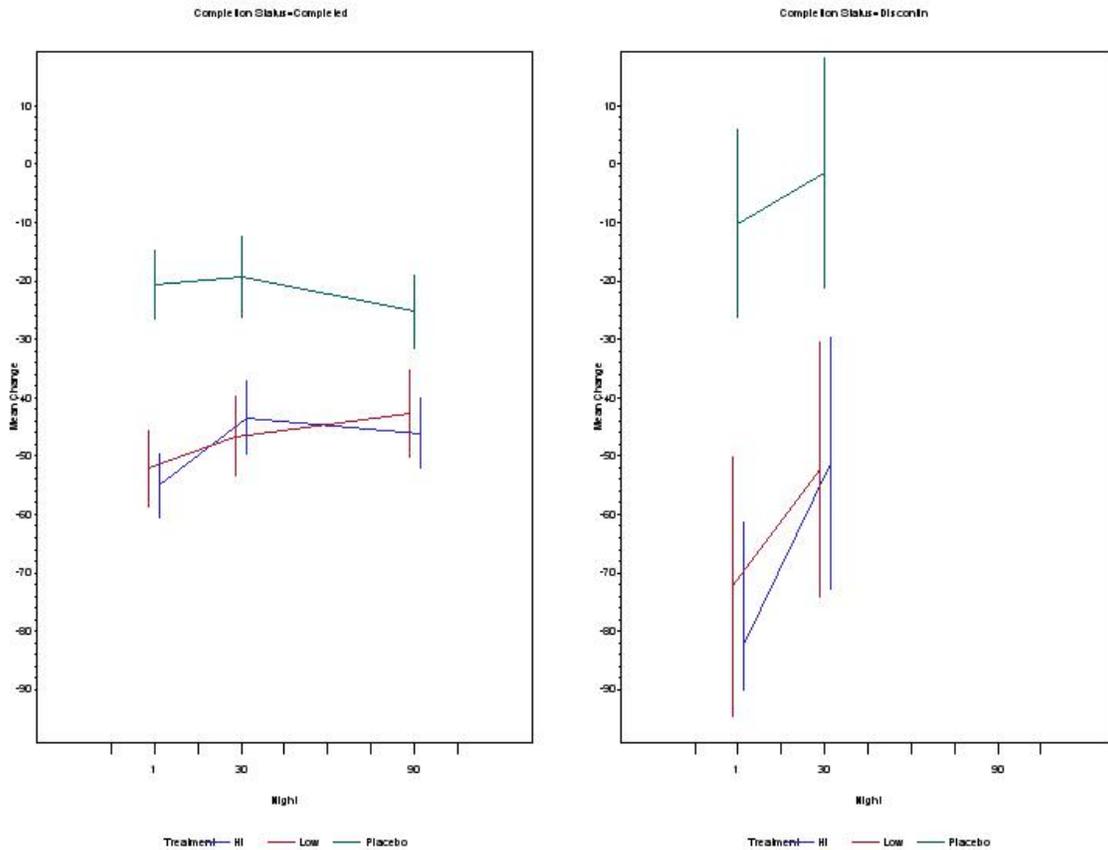
Figure 6 Study 28 Mean Change from Baseline in sTSO over Time by Completion Status



Note: N=42 placebo, 23 low dose, and 32 high dose discontinued

For objective WASO as for sTSO earlier effects were bigger for both MK4305 groups in the discontinued subgroup as seen in Figure 7. However, additional sensitivity analyses based on BOCF as well as a multiple imputation analysis using a placebo based model for all group's missing data supported the primary analysis. An analysis of completers (N=652) found an estimated mean difference from placebo of -25.7 for the high dose and -25.5 for the low dose at 1 month (both  $p < 0.0001$ ).

Figure 7 Study 28 Mean Change from Baseline in WASO over Time by Completion Status



Note: 39 placebo, 21 low dose, and 40 high dose discontinued

### 3.1.2 Study P029

The primary therapy period for the study was 28 July 2010 to 26 Oct 2011. The original protocol was dated February 8, 2010; there was a protocol clarification letter dated May 5, 2010.

#### 3.2.1.1 Study Design and Statistical Methods

This study design was nearly identical to that for study 28(so please refer to section 3.1.1.1 for details).

#### 3.2.1.2 Patient Disposition

Of the 1019 patients randomized into the study, 240 and 392 patients were randomized to suvorexant LD and HD, respectively and 387 were randomized to placebo. Of these, 10 patients were randomized but did not take any assigned study therapy (1, 5, and 4 patients in suvorexant LD, suvorexant HD, and placebo treatment groups respectively). Nine of these patients did not

meet entry criteria and were discontinued due to protocol violation prior to receiving study medication; one patient was accidentally randomized after the primary investigator withdrew from the study. Therefore, a total of 1009 patients received at least one dose of study drug and were evaluated for safety.

Of the treated subjects, a total of 881 (86.5%) patients completed the Treatment Phase, and 128 (12.6%) patients discontinued at some point during the study (see Table 9). The most common reason for discontinuing the study was due to an AE. The proportion of patients who discontinued during the Treatment Phase was similar between placebo (13.7%) and the suvorexant LD (14.2%) treatment groups. Fewer patients assigned to suvorexant HD group discontinued during treatment than placebo (10.5% vs 13.7%, respectively). The rate of discontinuation due to an AE was similar among treatment groups. The discontinuation rate among treatment groups was similar for non-elderly and elderly patients. The most common reason for discontinuing the study in both age cohorts was due to an AE.

**Table 9 Study 29: Disposition of Patients**

	MK-4305 LD n (%)	MK-4305 HD n (%)	Placebo n (%)	Total n (%)
<b>Not Randomized</b>				
Patients in population	240	392	387	1019
<b>Study Disposition</b>				
Not Treated	1 (0.4)	5 (1.3)	4 (1.0)	10 (1.0)
Completed Treatment†	205 (85.4)	346 (88.3)	330 (85.3)	881 (86.5)
Discontinued during Treatment	34 (14.2)	41 (10.5)	53 (13.7)	128 (12.6)
Adverse Event	10 (4.2)	19 (4.8)	17 (4.4)	46 (4.5)
Withdrawal by Subject	8 (3.3)	9 (2.3)	19 (4.9)	36 (3.5)
Protocol Violation	5 (2.1)	4 (1.0)	8 (2.1)	17 (1.7)
Lost to Follow-up	2 (0.8)	4 (1.0)	1 (0.3)	7 (0.7)
Lack of Efficacy	7 (2.9)	4 (1.0)	8 (2.1)	19 (1.9)
Physician Decision	2 (0.8)	1 (0.3)	0 (0.0)	3 (0.3)
<b>Protocol Milestone</b>				
Completed Treatment	205 (85.4)	346 (88.3)	330 (85.3)	881 (86.5)
Continuing Into Run-Out	205 (85.4)	344 (87.8)	327 (84.5)	876 (86.0)
Not Continuing Into Run-Out†	0 (0.0)	2 (0.5)	3 (0.8)	5 (0.5)
Discontinued during Treatment (not continuing into RO)	34 (14.2)	41 (10.5)	53 (13.7)	128 (12.6)
Each patient is counted once for Study Disposition and Protocol Milestone based on the latest corresponding disposition record.				
† AN 12041 attended V7 but did not complete all assessments.				
MK-4305 LD = MK-4305 20 mg for patients < 65 years and MK-4305 15 mg for patients ≥ 65 years.				
MK-4305 HD = MK-4305 40 mg for patients < 65 years and MK-4305 30 mg for patients ≥ 65 years.				

Note: This table was copied from page 159 of the sponsor's study report

### 3.2.1.1 Patient Demographics and Baseline Disease Characteristics

The ages ranged from 18 to 86 years with the mean age of 56 years. Of the treated patients, 410 (40.6%) were elderly. The distribution of age was similar among the treatment groups. Approximately two-thirds of patients enrolled in this study were female; further, the gender

distribution across treatment groups was also generally similar. Patient characteristics show that patient characteristics for the non-elderly and elderly were generally similar. Some differences between the two cohorts were observed in terms of region, race and ethnicity. In contrast to the PQ-Cohort, where most patients were recruited either in NA or Europe, in the Q-Cohort, most patients were from regions characterized as Other (which included Latin America, Korea, and India). The majority of the patients in the PQ-Cohort were White (91.3%) and Non-Hispanic or Latino (81.2%). On the other hand, only 48.7% of the patients in the Q-Cohort were White, 23.6% were Asian, 27% were Other, and approximately two third of the patients were Non-Hispanic or Latino. Other baseline characteristics (age, gender, BMI) were generally similar in the two cohorts.

**Table 10 Study 29: Baseline Demographics and Disease Characteristics (Randomized Patients)**

Variable	Statistic or Subgroup [N(%)]	Placebo	LD	HD	All
age	Mean (SD)	56.5 (15.4)	55.6 (16.1)	56.6 (15.0)	56.3 (15.4)
agegrp	< 65	231 (59.4)	144 (60.0)	233 (59.4)	608 (59.6)
agegrp	≥ 65	158 (40.6)	96 (40.0)	159 (40.6)	413 (40.4)
bmi	Mean (SD)	26.1 (4.1)	25.6 (3.9)	26.4 (4.3)	26.1 (4.2)
ethgrp	Hispanic	87 (22.4)	55 (22.9)	91 (23.2)	233 (22.8)
ethgrp	NonHispanic	302 (77.6)	185 (77.1)	301 (76.8)	788 (77.2)
race	AMERICAN INDIAN OR ALASKA NATI	2 (0.5)	.	1 (0.3)	3 (0.3)
race	ASIAN	27 (6.9)	27 (11.3)	28 (7.1)	82 (8.0)
race	BLACK	22 (5.7)	4 (1.7)	20 (5.1)	46 (4.5)
race	MULTIRACIAL	28 (7.2)	19 (7.9)	30 (7.7)	77 (7.5)
race	WHITE	310 (79.7)	190 (79.2)	313 (79.8)	813 (79.6)
sex	Female	250 (64.3)	158 (65.8)	271 (69.1)	679 (66.5)
sex	Male	139 (35.7)	82 (34.2)	121 (30.9)	342 (33.5)
cohort	Q	88 (22.6)	90 (37.5)	90 (23.0)	268 (26.2)
cohort	PQ	301 (77.4)	150 (62.5)	302 (77.0)	753 (73.8)
reggrp2	Asia/ Central and Eastern Europe	56 (14.4)	41 (17.1)	43 (11.0)	140 (13.7)
reggrp2	Central and South America	29 (7.5)	24 (10.0)	32 (8.2)	85 (8.3)
reggrp2	Europe	114 (29.3)	77 (32.1)	115 (29.3)	306 (30.0)
reggrp2	North America	190 (48.8)	98 (40.8)	202 (51.5)	490 (48.0)
LPS	Mean (SD)	68.0 (42.8)	65.3 (47.8)	67.3 (48.8)	67.2 (46.2)
TST	Mean (SD)	302.6 (64.0)	301.8 (68.2)	302.3 (67.5)	302.3 (66.2)
WASO	Mean (SD)	118.4 (49.1)	119.6 (50.8)	119.4 (51.3)	119.0 (50.3)
sTSTm	Mean (SD)	310.2 (77.1)	298.3 (81.9)	315.3 (77.0)	309.3 (78.5)

### 3.1.2.1 Sponsor's Results

Based upon findings from a monitoring visit, an audit was performed on Site 120 from Russia which suggested that data from this site may lack reliability and integrity. Therefore, in a protocol clarification letter the sponsor stipulated that this site was not to be included in the primary efficacy analysis.

Based on the multiplicity testing strategy:

Suvorexant HD (40 mg for patients <65 years and 30 mg for patients  $\geq$ 65 years) was superior to placebo in improving sleep maintenance as measured by the change from baseline in sTSTm and change from baseline in WASO, at Month 1, Month 3, and Week 1/Night 1, respectively.

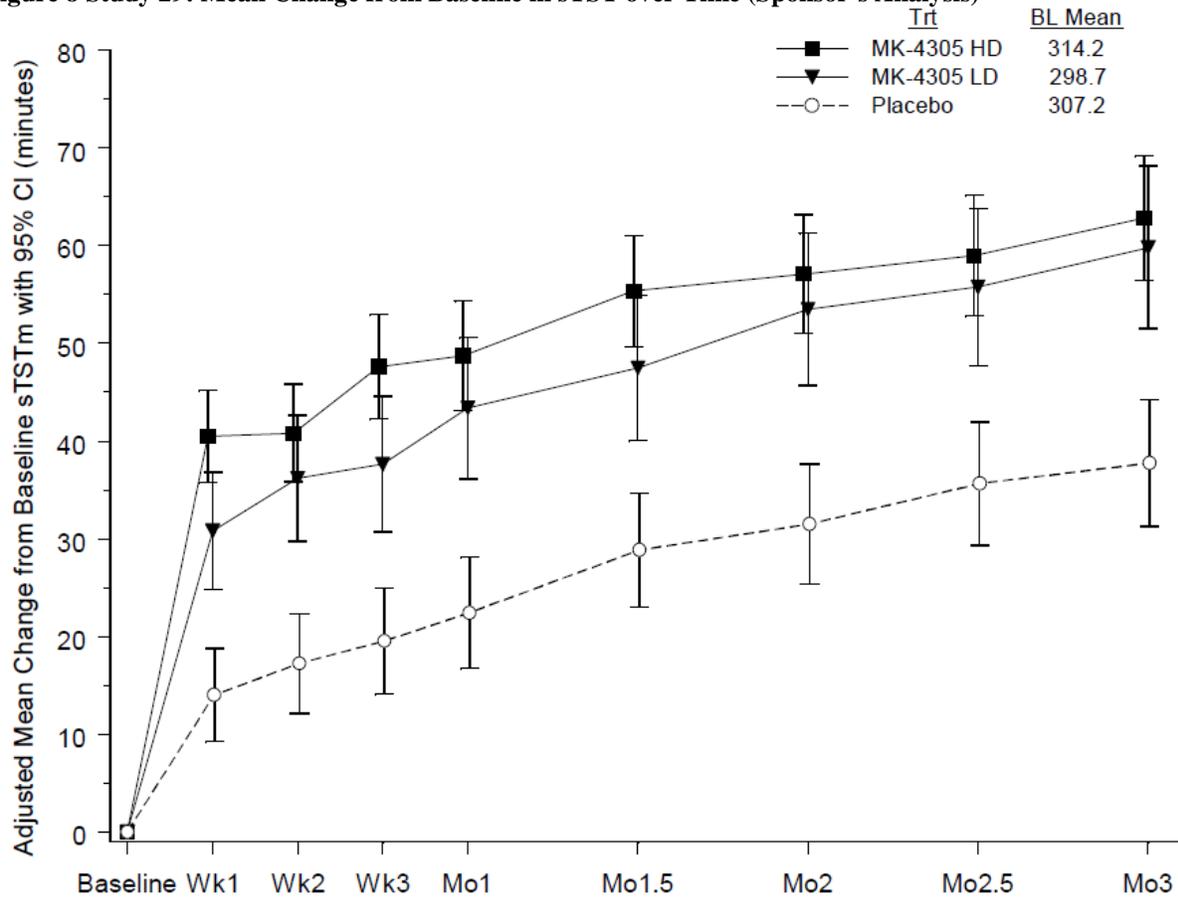
Suvorexant HD (40 mg for patients <65 years and 30 mg for patients  $\geq$ 65 years) was superior to placebo in improving sleep onset as measured by the change from baseline in sTSOm, at Month 1, Month 3, and Week 1/Night 1, respectively. It was also superior to placebo in improving sleep onset as measured by the change from baseline in LPS, at Month 1 and Night 1, respectively, but not at Month 3.

#### **Subjective Sleep Maintenance –sTSTm**

Figure 8 and Table 11 present summary statistics and analysis results for sTSTm at the key timepoints during the Treatment Phase for both the HD and LD of suvorexant. The analyses suggest that suvorexant HD was superior to placebo in increasing subjective total sleep time at the primary (Month 1 and Month 3) timepoints and also at the secondary (Week 1) timepoint (all p-values < 0.00001). The increases from baseline in sTSTm for suvorexant HD ranged from 40.4 to 62.8 minutes on average and were 25.1 to 26.4 minutes greater than those for placebo.

The increases in sTSTm from baseline for suvorexant LD ranged from 30.8 to 59.8 minutes on average and were 16.8 to 22.1 minutes greater than those for placebo.

Figure 8 Study 29: Mean Change from Baseline in sTST over Time (Sponsor's Analysis)



Note: This figure was copied from page 205 of the sponsor's study report

**Table 11 Study 29: Change from Baseline in Mean Subjective Total Sleep Time (Sponsor's Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) <sup>†</sup>
<b>Week 1</b>					
MK-4305 LD	231	298.7 ( 82.2)	331.0 ( 86.9)	32.3 ( 50.8)	30.8 ( 24.8, 36.8)
MK-4305 HD	373	314.2 ( 77.4)	354.2 ( 83.1)	39.9 ( 51.1)	40.4 ( 35.7, 45.1)
Placebo	364	307.2 ( 77.2)	322.1 ( 81.0)	14.9 ( 39.6)	14.0 ( 9.2, 18.7)
<b>Month 1</b>					
MK-4305 LD	219	300.8 ( 78.8)	346.1 ( 90.0)	45.3 ( 60.5)	43.4 ( 36.1, 50.6)
MK-4305 HD	365	314.0 ( 77.4)	362.2 ( 81.5)	48.2 ( 60.6)	48.7 ( 43.1, 54.3)
Placebo	350	308.2 ( 77.6)	332.3 ( 80.8)	24.1 ( 51.0)	22.4 ( 16.7, 28.1)
<b>Month 3</b>					
MK-4305 LD	197	306.7 ( 69.5)	371.6 ( 72.3)	64.8 ( 62.7)	59.8 ( 51.5, 68.1)
MK-4305 HD	340	314.8 ( 75.9)	378.3 ( 79.3)	63.5 ( 67.3)	62.8 ( 56.4, 69.2)
Placebo	325	311.8 ( 71.8)	352.7 ( 78.4)	40.9 ( 60.7)	37.7 ( 31.2, 44.2)
<b>Pairwise Comparison</b>		<b>Difference in LS Means (95% CI)<sup>†</sup></b>			<b>p-Value<sup>†</sup></b>
<b>Week 1</b>					
MK-4305 HD vs. Placebo		26.4 ( 19.8, 33.1)			<0.00001
MK-4305 LD vs. Placebo		16.8 ( 9.1, 24.5)			0.00002
<b>Month 1</b>					
MK-4305 HD vs. Placebo		26.3 ( 18.3, 34.3)			<0.00001
MK-4305 LD vs. Placebo		20.9 ( 11.7, 30.2)			<0.00001
<b>Month 3</b>					
MK-4305 HD vs. Placebo		25.1 ( 16.0, 34.2)			<0.00001
MK-4305 LD vs. Placebo		22.1 ( 11.5, 32.6)			0.00004
<sup>†</sup> Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, cohort, gender, treatment, time point, and treatment-by-time point interaction as covariates. MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years. MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 204 of the sponsor's study report

### Objective Sleep Maintenance – WASO

Table 12 presents summary statistics and analysis results for WASO at the key timepoints during the Treatment Phase for both suvorexant HD and LD. The analyses show that suvorexant HD was superior to placebo in decreasing WASO at the primary (Month 1 and Month 3) timepoints and also at the secondary (Night 1) timepoint (all p-values <0.0001). The decreases from baseline in WASO for suvorexant HD ranged from 51.9 to 63.3 minutes on average and were 29.4 to 42.0 minutes greater than those for placebo.

Nominal p-values also suggest that suvorexant LD was more efficacious than placebo in decreasing WASO at the primary and secondary timepoints (all nominal p-values <0.0001). The decreases from baseline in WASO for suvorexant LD ranged from 46.6 to 58.3 minutes on average and were 24.1 to 37.0 minutes greater than those for placebo.

**Table 12 Study 29: Change from Baseline in Mean Objective WASO (Sponsor's Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) <sup>†</sup>
<b>Night 1</b>					
MK-4305 LD	144	119.3 (50.8)	61.3 (41.5)	-58.0 (49.9)	-58.3 (-65.0, -51.6)
MK-4305 HD	285	119.5 (51.5)	55.9 (37.5)	-63.5 (51.6)	-63.3 (-68.0, -58.6)
Placebo	283	118.3 (49.4)	97.9 (54.5)	-20.4 (57.5)	-21.3 (-26.1, -16.6)
<b>Month 1</b>					
MK-4305 LD	132	118.0 (51.7)	72.5 (48.9)	-45.4 (56.5)	-46.6 (-53.8, -39.3)
MK-4305 HD	278	120.9 (51.2)	67.3 (40.8)	-53.5 (53.5)	-51.9 (-56.9, -46.9)
Placebo	270	118.0 (49.3)	96.2 (56.4)	-21.8 (56.7)	-22.5 (-27.5, -17.4)
<b>Month 3</b>					
MK-4305 LD	127	120.1 (52.5)	64.7 (43.4)	-55.5 (53.8)	-56.0 (-63.3, -48.6)
MK-4305 HD	260	121.6 (51.1)	65.5 (40.2)	-56.1 (51.4)	-54.2 (-59.3, -49.1)
Placebo	252	118.3 (49.6)	94.1 (55.1)	-24.1 (59.7)	-24.8 (-30.0, -19.6)
Pairwise Comparison		Difference in LS Means (95% CI) <sup>†</sup>		p-Value <sup>†</sup>	
<b>Night 1</b>					
MK-4305 HD vs. Placebo		-42.0 (-48.6, -35.3)		<.00001	
MK-4305 LD vs. Placebo		-37.0 (-45.1, -28.8)		<.00001	
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-29.4 (-36.6, -22.3)		<.00001	
MK-4305 LD vs. Placebo		-24.1 (-33.0, -15.3)		<.00001	
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-29.4 (-36.7, -22.1)		<.00001	
MK-4305 LD vs. Placebo		-31.1 (-40.1, -22.2)		<.00001	
<sup>†</sup> Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates. MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years. MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: this table was copied from page 210 of the sponsor's study report

### Subjective Sleep Onset – sTSOm

Table 13 presents summary statistics and analysis results for sTSOm at the key timepoints during the Treatment Phase for both suvorexant HD and LD. The analyses show that suvorexant HD was superior to placebo in decreasing sTSOm at the primary (Month 1 and Month 3) timepoints and also at the secondary (Week 1) timepoint (all p-values <0.00005). The decreases from baseline in sTSOm for suvorexant HD ranged from 19.7 to 33.7 minutes on average and were 12.8 to 13.2 minutes greater than those for placebo.

Nominal p-values also suggest that suvorexant LD was more efficacious than placebo in decreasing sTSOm at the primary and secondary timepoints (all nominal p-values <0.05000). The decreases from baseline in sTSOm for suvorexant LD ranged from 14.2 to 28.1 minutes on average and were 6.9 to 7.6 minutes greater than those for placebo.

**Table 13 Study 29 Mean Change from Baseline in sTSO (Sponsor’s Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) <sup>†</sup>
<b>Week 1</b>					
MK-4305 LD	231	86.1 ( 78.3)	70.6 ( 75.9)	-15.5 ( 35.3)	-14.2 (-18.4, -10.0)
MK-4305 HD	373	75.2 ( 62.5)	56.1 ( 59.0)	-19.1 ( 33.2)	-19.7 (-23.0, -16.4)
Placebo	364	82.6 ( 77.5)	74.7 ( 76.0)	-7.9 ( 32.9)	-6.7 (-10.0, -3.3)
<b>Month 1</b>					
MK-4305 LD	219	83.1 ( 74.5)	60.4 ( 74.6)	-22.7 ( 47.9)	-21.0 (-26.4, -15.6)
MK-4305 HD	365	75.1 ( 62.8)	48.7 ( 54.2)	-26.5 ( 46.2)	-26.9 (-31.1, -22.8)
Placebo	350	82.0 ( 77.8)	66.5 ( 66.8)	-15.5 ( 43.6)	-14.1 (-18.4, -9.9)
<b>Month 3</b>					
MK-4305 LD	197	75.5 ( 54.8)	43.7 ( 36.8)	-31.7 ( 44.7)	-28.1 (-33.7, -22.4)
MK-4305 HD	340	74.1 ( 62.4)	40.3 ( 38.6)	-33.8 ( 50.7)	-33.7 (-38.0, -29.3)
Placebo	325	77.2 ( 66.1)	55.4 ( 56.0)	-21.8 ( 47.5)	-20.5 (-24.9, -16.1)
Pairwise Comparison		Difference in LS Means (95% CI) <sup>†</sup>			p-Value <sup>†</sup>
<b>Week 1</b>					
MK-4305 HD vs. Placebo		-13.1 (-17.7, -8.4)			<0.00001
MK-4305 LD vs. Placebo		-7.5 (-12.9, -2.2)			0.00593
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-12.8 (-18.8, -6.9)			0.00003
MK-4305 LD vs. Placebo		-6.9 (-13.7, -0.0)			0.04975
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-13.2 (-19.4, -7.0)			0.00003
MK-4305 LD vs. Placebo		-7.6 (-14.7, -0.4)			0.03894
<sup>†</sup> Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, cohort, gender, treatment, time point, and treatment-by-time point interaction as covariates. MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years. MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 213 of the sponsor’s study report

### Objective Sleep Onset – LPS

Table 14 and Figure 9 present summary statistics and analysis results for LPS at the key timepoints during the Treatment Phase for both suvorexant HD and LD. The analyses show that suvorexant HD was superior to placebo in decreasing LPS at the primary Month 1 timepoint and the secondary Night 1 timepoint (p-values <0.00005), but not at the primary Month 3 timepoint (p-value = 0.26510). The decreases from baseline in LPS for suvorexant HD ranged from 34.7 to 36.7 minutes on average and were 12.7 to 12.1 minutes greater than those for placebo at Night 1 and Month 1, respectively. While the decrease in LPS from baseline at Month 3 of 32.2 minutes for suvorexant HD was nearly as large as the differences observed at Night 1 and Month 1, indicative of a sustained response to suvorexant, the increasing placebo response between Night 1 and Month 3 may have contributed to a non-significant difference from placebo of only 3.6 minutes (p-value = 0.26510).

Nominal p-values also suggest that suvorexant LD was more efficacious than placebo in decreasing LPS at the primary Month 1 and secondary Night 1 timepoints (nominal p-values

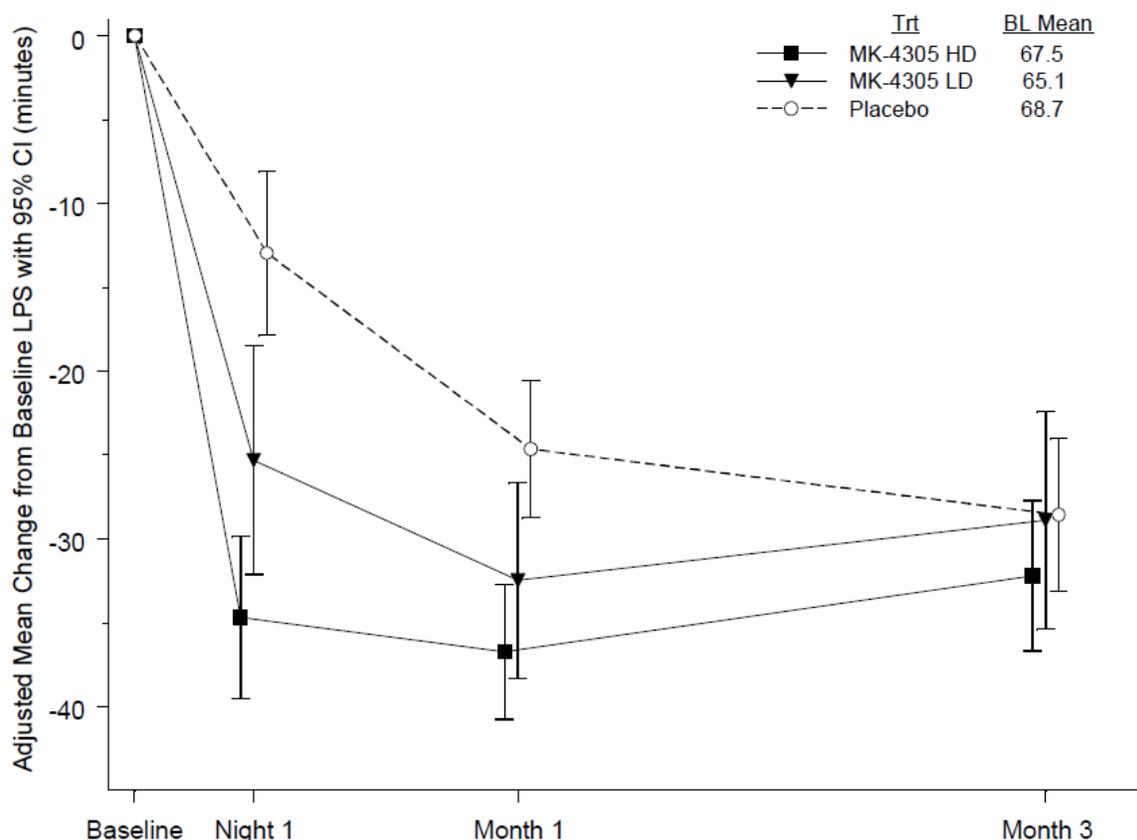
<0.05000). The decreases from baseline in LPS for suvorexant LD ranged from 25.3 to 32.5 minutes on average and were 12.4 and 7.8 minutes greater than those for placebo at Night 1 and Month 1, respectively. The decrease from baseline in LPS at Month 3 of 28.9 minutes for suvorexant LD was in fact greater than the change from baseline of 25.3 observed at Night 1, but the increasing placebo response between Night 1 and Month 3 resulted in a difference from placebo of only 0.3 minutes which had a nominal p-value = 0.93219.

**Table 14 Study 29: Mean Change from Baseline in LPS (Sponsor's Analysis)**

Treatment	N	Baseline Mean (SD)	Time Point Mean (SD)	Change from Baseline at Time Point	
				Mean (SD)	LS Mean (95% CI) <sup>†</sup>
<b>Night 1</b>					
MK-4305 LD	144	65.1 (47.0)	41.9 (32.7)	-23.1 (40.3)	-25.3 (-32.2, -18.5)
MK-4305 HD	289	67.5 (49.4)	33.0 (24.2)	-34.5 (48.0)	-34.7 (-39.5, -29.9)
Placebo	284	68.7 (43.3)	54.9 (60.8)	-13.8 (62.3)	-13.0 (-17.8, -8.1)
<b>Month 1</b>					
MK-4305 LD	133	66.2 (47.4)	34.6 (26.6)	-31.6 (47.5)	-32.5 (-38.3, -26.7)
MK-4305 HD	280	66.7 (46.5)	30.8 (28.2)	-35.9 (48.3)	-36.7 (-40.8, -32.7)
Placebo	271	68.9 (43.3)	43.2 (44.4)	-25.7 (53.9)	-24.6 (-28.7, -20.6)
<b>Month 3</b>					
MK-4305 LD	127	66.1 (48.3)	38.0 (34.9)	-28.1 (47.8)	-28.9 (-35.4, -22.4)
MK-4305 HD	262	67.3 (47.6)	35.6 (42.6)	-31.8 (55.3)	-32.2 (-36.7, -27.7)
Placebo	255	70.2 (44.0)	39.6 (36.1)	-30.6 (49.5)	-28.6 (-33.1, -24.0)
<b>Pairwise Comparison</b>		<b>Difference in LS Means (95% CI)<sup>†</sup></b>		<b>p-Value<sup>†</sup></b>	
<b>Night 1</b>					
MK-4305 HD vs. Placebo		-21.7 (-28.6, -14.9)		<.00001	
MK-4305 LD vs. Placebo		-12.4 (-20.7, -4.0)		0.00392	
<b>Month 1</b>					
MK-4305 HD vs. Placebo		-12.1 (-17.8, -6.4)		0.00004	
MK-4305 LD vs. Placebo		-7.8 (-15.0, -0.7)		0.03063	
<b>Month 3</b>					
MK-4305 HD vs. Placebo		-3.6 (-10.1, 2.8)		0.26510	
MK-4305 LD vs. Placebo		-0.3 (-8.3, 7.6)		0.93219	
<sup>†</sup> Based on a mixed effects model with terms for baseline value, age category (<65, ≥65), region, gender, treatment, time point, and treatment-by-time point interaction as covariates.					
MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years.					
MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.					

Note: This table was copied from page 216 of the sponsor's study report

**Figure 9 Study 29: Adjusted Means for Change from Baseline in Latency to Persistent Sleep**



Note: This figure was copied from page 217 of the sponsor's study report

### Additional Sponsor Sensitivity Analyses

To evaluate the robustness of these objective sleep onset results, a nonparametric procedure using aligned ranks on change from baseline in LPS was performed. This procedure confirmed statistical significance of the reduction in LPS for suvorexant HD at Night 1, but missed statistical significance at Month 1 (the p-value of 0.03874 was not  $\leq 0.02500$ ). The aligned rank procedure performed at Month 3 confirmed the non-significant result of the primary analysis approach for Month 3 (p-value = 0.14313). The nominal p-values for the comparison of suvorexant LD versus placebo using the aligned rank procedure were  $> 0.05000$  for each of the 3 key timepoints. Since the results of the analysis at Month 1 using the aligned rank procedure provided a different conclusion than the primary mixed model analysis approach for Month 1, another rank procedure (ETRANK) was performed on LPS at Month 1 as specified by the data analysis plan in the protocol. Note that this method does not involve imputation, using only observed data. The results using the ETRANK procedure show that LPS was significantly reduced for suvorexant HD as compared with placebo at Month 1. ETRANK penalizes the ranks of patients who dropped out due to lack of efficacy or AE (i.e., drop-outs potentially related to treatment), and it also uses Entsuah scores that are symmetric, which may contribute to the

difference in results given by these two rank procedures. The ETRANK procedure performed on LPS at Month 3 also showed a significant reduction in LPS for suvorexant HD as compared with placebo at Month 3.

*Reviewer's Comment: This reviewer did not verify the ETRANK sensitivity analysis since it is not implemented in SAS, no statistical program was provided by the sponsor, and it is only a secondary sensitivity analysis.*

To further explore the comparison of suvorexant and placebo with respect to LPS, a post-hoc analysis was also performed on changes from baseline in log-transformed LPS, log(LPS). The log-transformation reduces the influence of potential outliers and departures from normality. The results of the analysis on log-transformed LPS agree with the results from the primary analysis for suvorexant HD for each of the key timepoints; LPS was significantly reduced for suvorexant HD as compared with placebo at Night 1 and Month 1, but not at Month 3. Comparisons of suvorexant LD and placebo with respect to changes from baseline in log-transformed LPS were not significant at the key timepoints (p-values > 0.05).

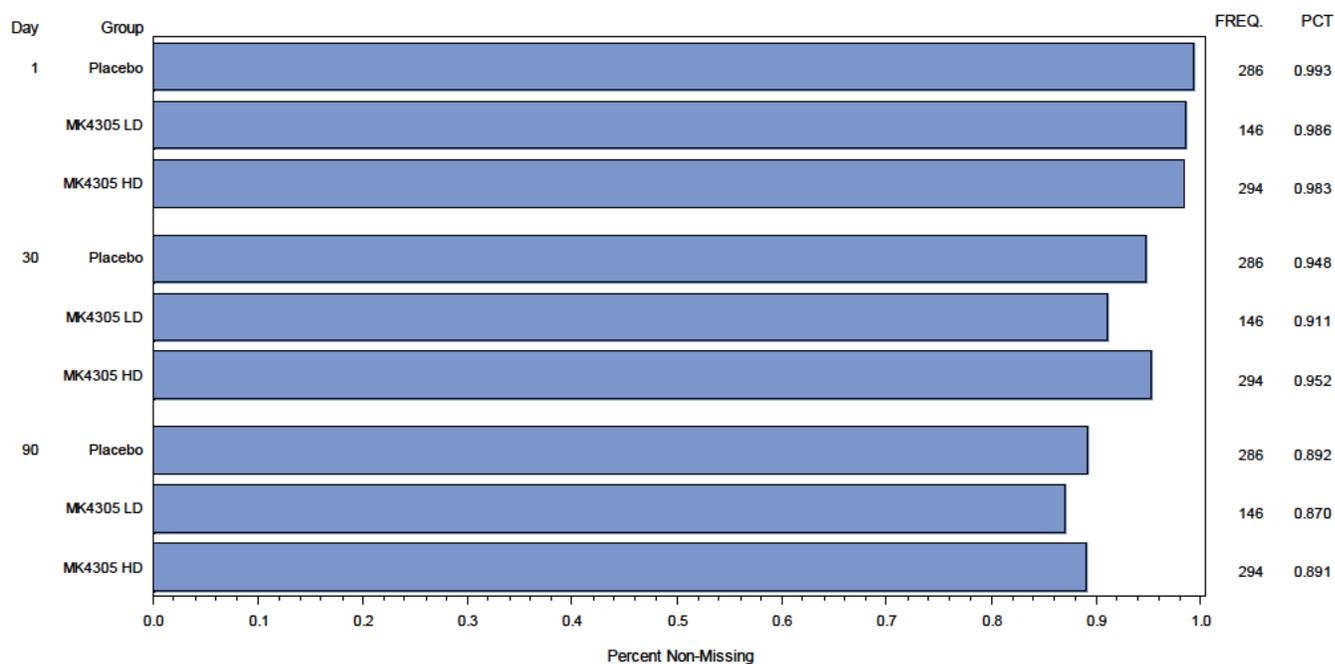
### 3.1.2.2 Reviewer's Results

There were three multiple enrollees: one enrolled twice within study 029, another sequentially between study 28 and 29 and another had overlapping enrollments in study 28 and 29. These were handled as prespecified in the protocol clarification letter and described in section 3.1.1.1 on page 8.

Because of the prespecified order of testing the various timepoints and the lack of significance of LPS at Month 3 which was to be tested before Night 1 we cannot conclude efficacy in terms of LPS or (subjective TSO) on Night 1. In addition, any secondary endpoint results assessing onset must from a multiplicity perspective be viewed as exploratory at best, because of the lack of significance for LPS at month 3.

Figure 10 shows the percentages of randomized patients with non-missing LPS data at each timepoint for each group. For the placebo and high dose groups between 5 and 6% had missing data at month 1 and about 11% had missing data at Month 3. The extent of missingness for the other primary efficacy measures was similar.

**Figure 10 Study 29: Percentages of Randomized Patients with non-Missing LPS data over Time**



An LOCF analysis of covariance for the Month 3 LPS gave  $p=0.0550$ , while an observed case analysis gave  $p=0.2838$ . So these additional analyses confirmed the sponsor's insignificant primary analysis result at Month 3 for LPS.

The somewhat strange phenomenon observed in study 28 of the low dose appearing markedly better in terms of sTSO in the subgroup of discontinued patients (Figure 6) compared to placebo or completers at times where they had data was not replicated in study 29 (nor was it seen for sTST or WASO either). Completers analyses of sTSO (N=862, Month 1 low diff: -9.7,  $p=0.0019$ , high diff: -12.9,  $p<0.0001$ ), sTST (Month 1 low diff: 22.6 and high diff: 24.9, both  $p<0.0001$ ) and WASO (N=647, Month 1 low diff: 23.6, high diff: 30.8, both  $p<0.0001$ ) supported the primary analyses. The following additional sensitivity analyses initiated by the reviewer also supported the primary analysis for sTST and objective WASO: baseline carried forward in the event of missing data analyses and multiple ( $m=20$ ) imputation based on a placebo group based imputation model applied to impute missing data for all groups.

The sponsor prespecified excluding site 45490 from the primary analysis before unblinding the data because of protocol deviations they had discovered at that site. For completeness it is important to look at the results including the patients from this site since they were randomized and had post-baseline data. As it turns out the efficacy conclusions are unaffected by the inclusion or exclusion of data from this site as shown in Table 15.

**Table 15 Study 29: Primary Efficacy Results for High Dose Including Site 45490**

<b>Endpoint</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>P-value</b>
<b>sTSTm</b>	<b>1</b>	26.43	3.40	<.0001
	<b>30</b>	26.22	4.07	<.0001
	<b>90</b>	25.06	4.62	<.0001
<b>WASO</b>	<b>1</b>	-41.61	3.47	<.0001
	<b>30</b>	-29.54	3.58	<.0001
	<b>90</b>	-29.56	3.73	<.0001
<b>sTSOm</b>	<b>1</b>	-13.01	2.37	<.0001
	<b>30</b>	-12.80	3.02	<.0001
	<b>90</b>	-13.19	3.14	<.0001
<b>LPS</b>	<b>1</b>	-22.01	3.68	<.0001
	<b>30</b>	-12.92	3.04	<.0001
	<b>90</b>	-4.51	3.37	0.1818

A sensitivity analysis with patients reporting more than 24 hours in a day's data unmodified (rather than excluded as prespecified) yielded the following results (Table 16). The exclusion of this questionable data did not seem to impact the results.

**Table 16 Study 29: Sensitivity Analysis for Primary Subjective Endpoints (High Dose vs. Placebo)**

<b>Endpoint</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>P-value</b>
<b>sTSOm</b>	<b>1</b>	-11.8	3.0	<.0001
	<b>30</b>	-12.7	3.1	<.0001
	<b>90</b>	-12.9	3.6	0.0003
<b>sTSTm</b>	<b>1</b>	26.0	3.8	<.0001
	<b>30</b>	26.1	4.3	<.0001
	<b>90</b>	25.1	5.0	<.0001

### 3.1.3 Study P006

#### 3.1.3.1 SUMMARY OF STUDY DESIGN

This is a randomized, double-blind (with in-house blinding), placebo-controlled, 2-period crossover PSG study to assess the safety, tolerability, and efficacy of four doses of MK-4305 (10, 20, 40, and 80 mg) in the treatment of patients with primary insomnia. This design is equivalent to four separate 2-treatment 2-period crossover trials with patients in each trial receiving one of the four MK-4305 doses and placebo. The objectives of the study include demonstration of the effectiveness and dose response of MK-4305 and identification of an appropriate dose or doses for Phase III development. An interim analysis (IA) was to be employed to assess: (1) preliminary effectiveness of MK-4305 (futility of doses), and (2) safety and tolerability. A second IA was to be performed to aid in planning and/or acceleration of future studies. Each interim analysis was to include patients from the ex-Japan stratum only.

During the MK-4305 treatment period, patients were to receive one of four possible MK-4305 doses (10, 20, 40, or 80 mg). Assignment to dose of MK-4305 and treatment sequence (MK-4305/Placebo or Placebo/MK-4305) was to be randomly determined by a computer-generated random allocation schedule. The randomization was to be stratified by geographic region (Japan vs. ex-Japan). Each 4-week treatment period was to consist of an overnight PSG visit on the first and last night of the treatment period, with an interim office visit at the Week-2 midpoint. Each treatment period was to consist of  $28 \pm 2$  days of treatment. A single-blind placebo washout interval of a minimum of 7 days was to separate the two treatment periods. Additional treatment sequences might have been added to evaluate a 5-mg dose pending the results of an interim futility analysis when approximately 50% of the patients had completed the study.

#### **Objectives:**

1. To evaluate the efficacy of MK-4305 compared with placebo in improving sleep efficiency (SE) as measured by polysomnography (PSG) on Night 1 and at the end of 4 weeks of treatment, where SE is defined as 100 times total sleep time (minutes) divided by time in bed (minutes).

An interim analysis for identifying futile doses will be conducted after 50% of patients have completed their two 4-week treatment periods. A standing internal data monitoring committee (siDMC) will evaluate these interim data, encompassing evaluation of efficacy data in conjunction with evaluation of other data (e.g., adverse event data), to determine what if any actions will be taken for the remainder of the study. Based upon this analysis, it is possible that a lower dose of MK-4305 (5 mg) may be added to the study; in this case, the sample size may increase by up to 52 patients (due to the addition of these lower dose sequences). A second interim analysis may be conducted when approximately 160 patients have completed both

periods of the crossover study to help expedite Phase III planning. Each interim analysis will include patients from the ex- Japan stratum only.

### **Primary and Secondary Endpoints**

The co-primary endpoints that will be used in the evaluation of the primary efficacy hypothesis are sleep efficiency (SE) measurements at Night 1 and at the end of 4 weeks of treatment. SE is defined as total sleep time (TST) in minutes divided by time in bed (measured from lights off to lights on; fixed at 8 hours on each PSG night) in minutes, multiplied by 100, where TST is defined as the total time (minutes) in Stages 1, 2, 3, 4 and REM.

The secondary endpoints include Night 1 and End of Week 4 measurements for: 1) wakefulness after sleep onset (WASO) and 2) latency to persistent sleep (LPS). WASO is defined as the duration of wakefulness (any epoch of Stage 0) from persistent sleep onset (first epoch of the first twenty consecutive epochs of non-wake) to lights on and LPS is defined as the duration of time from lights off to persistent sleep onset. An epoch of non-wake is defined as a 30-second interval classified as either Stage 1, 2, 3, 4 or REM according to conventional Rechtschaffen and Kales (R&K) scoring.

The Full Analysis Set (FAS) population will serve as the primary population for the analysis of efficacy data in this study. The FAS population is a subset of all randomized patients with patients excluded for the following reasons: failure to receive at least one dose of study treatment lack of any post-randomization efficacy data subsequent to at least one dose of study treatment.

The primary and secondary hypotheses will be evaluated using a mixed effects model including terms for baseline value, geographic region (Japan vs. ex-Japan), treatment, sequence, period, time (as a categorical variable), and treatment-by-time and period-by-time interactions. An unstructured covariance matrix will be used. If the model does not converge then a random effect for patient will be added to the model and an unstructured 2 x 2 covariance matrix within patient and time will be used. Note that the baseline value is only measured prior to randomization to a treatment sequence. The model will be used to provide an estimate of the treatment effect for the comparison of each MK-4305 dose with placebo.

Multiple imputation will be also used as a sensitivity analysis to approximate the missing value for the primary, secondary (and selected exploratory) endpoints.

### **Multiplicity**

An interim analysis for futility was to be conducted when approximately 50% of the patients had completed both treatment periods for the study; this analysis would ideally not affect the Type I error. A second interim analysis was to possibly be conducted when approximately 160 patients completed both periods of the crossover study for planning future Phase III studies. If this analysis was performed, the method of Haybittle-Peto was to be used to control overall Type I error; this method uses an alpha of 0.001 at the interim and an alpha of 0.05 at the end of the study. The interim analyses were to include patients from the ex-Japan stratum only. While, nominal p-values were to be computed for all comparisons of MK-4305 with placebo, statistical significance for the primary and secondary hypotheses was to be based on the following multiplicity strategy. To account for the multiple dose comparisons to placebo for the primary

efficacy hypothesis, a fixed sequential testing procedure was to be used to assess statistical significance (for all doses not deemed futile at the interim analysis), beginning with the highest dose (i.e., 80 mg). The comparison of MK-4305 with placebo for Sleep Efficiency (SE) had to be significant at  $\alpha=0.05$  at both time points (i.e., Night 1 and Week 4) in order to assess the statistical significance of the comparison of MK-4305 with placebo for the next highest dose (i.e., 40 mg), and so on. If a non-significant result was observed at either time point, the differences between MK-4305 and placebo for SE were to be considered nonsignificant at this and all lower doses. This procedure was to provide strong control for the multiple dose comparisons of MK-4305 with placebo for the primary efficacy endpoint. MK-4305 doses that were statistically significant for the primary hypothesis were to be tested in a similar fashion for the first secondary endpoint (WASO) at the 5% level of significance. Similarly, MK-4035 doses which are statistically significantly different from placebo for SE and WASO at both time points were to be tested in a similar fashion for the second secondary endpoint (LPS). This would provide control (not strong) of alpha for multiple dose comparisons of MK-4305 to placebo for multiple secondary endpoints.

### **Sample Size Determination**

Mean improvement in SE, WASO and LPS after 1 night of treatment are available from Actelion's orexin compound, almorexant. A total of 208 patients (i.e., 52 for each of the 4 doses) completing both periods of the study, including ~ 40 patients from Japanese ancestry, will be required to compare SE between all four MK-305 doses and placebo. Based upon this sample size, there is approximately 95% (73%) power to detect a difference of 8.33 (6.25) in SE at both time points for a particular MK-4305 dose. (Note that a difference of 8.33 (6.25) in SE corresponds to a 40 (30) minute difference in total sleep time (TST) when time in bed is fixed at 8 hours.) It is expected that approximately 10% of the patients randomized will discontinue treatment permanently during one of the two 4-week treatment periods; therefore, approximately 250 patients will be randomized to ensure 208 with complete data.

### **Interim Analyses**

An interim analysis for futility was to be conducted after 50% of patients had completed their two 4-week treatment periods. The following guidelines were to be used by a standing internal data monitoring committee (siDMC), encompassing evaluation of efficacy data in conjunction with evaluation of other data (e.g., adverse event data), to determine what if any actions were to be taken for the remainder of the study. If the conditional probability of finding a significant difference from placebo for SE was  $<20\%$  at either Night 1 or at the end of Week 4 for the lowest two doses, future enrollment may have been stopped for the lower of the two doses. A similar analysis was to be performed for each higher MK-4305 dose pair, in order of increasing doses. If the conditional probability of finding a significant difference for a dose was  $20\%$  at Night 1 and at the end of Week 4, then the study was to continue with this dose pair and all higher doses unless the siDMC decided otherwise on the basis of safety and/or tolerability issues. Guidelines regarding the termination of a dose on the basis of safety and/or tolerability issues were to be provided in the siDMC charter. An MK-4305 dose which had  $20\%$  conditional probability of a significant difference in SE at both Night 1 and at the end of Week 4 but had a poor profile for safety and/or tolerability as compared with placebo was to not be continued. If, based upon the siDMC's review, the lower dose of MK-4305 (5 mg) was recommended for evaluation (e.g., if none of the MK-4305 doses were futile and/or there was poor tolerability) this dose may have been included for the remainder of the study. Such a decision could have

increased the total patients completing the study by 26 to 52, depending upon the siDMC's recommendation. .

A second interim analysis was to potentially be conducted when approximately 160 patients had completed both periods of the crossover study. This analysis was to include patients in the ex-Japan stratum only. The purpose of this analysis was to expedite planning for Phase III studies; therefore both efficacy and safety were to be evaluated. The study was not to be stopped for superior efficacy on the basis of this analysis; however, the method of Haybittle-Peto was to be used to account for multiplicity related to this evaluation.

### **3.1.3.2 Patient Disposition**

Out of the 254 randomized patients, 249 received at least 1 dose of placebo and 243 received at least 1 dose of MK-4305. A total of 228 patients completed the study and 26 patients discontinued from the study (23 patients discontinued the study during Treatment Period 1 or 2 and 3 patients discontinued the study during the Washout period). The proportion of patients who discontinued during the treatment periods were similar between placebo and the MK-4305 total group. There were 3 patients who discontinued therapy in Treatment Period 1 but continued in the study by entering the Washout period (as allowed per protocol).

### **3.1.3.3 Patient Demographics**

The ages range from 18 to 64 years with the mean age of 44 years. The distribution of age was similar among the treatments and treatment sequences. A greater proportion of females were enrolled in this study; the total percent of females was 58.3%. While the gender distribution tended to range a bit for the treatment sequences, the distribution was similar for the treatments groups (ranging from 54.2% to 65.6% female).

**Table 17 Study 06: Baseline Demographics by Sequence**

Arm	Variable	Statistic or Sub Category	Statistic
10 mg: Placebo	age	Mean (SD)	44.5 (11.2)
Placebo: 10 mg	age	Mean (SD)	46.0 (11.9)
20 mg: Placebo	age	Mean (SD)	43.3 (11.4)
Placebo: 20 mg	age	Mean (SD)	44.4 (11.1)
30 mg: Placebo	age	Mean (SD)	45.8 (10.0)
Placebo: 30 mg	age	Mean (SD)	43.7 (12.6)
40 mg: Placebo	age	Mean (SD)	43.1 (12.0)
Placebo: 40 mg	age	Mean (SD)	44.1 (12.3)
All	age	Mean (SD)	44.4 (11.5)
10 mg: Placebo	race	ASIAN N(%)	4 (12.9)
Placebo: 10 mg	race	ASIAN	7 (21.9)
20 mg: Placebo	race	ASIAN	5 (15.2)
Placebo: 20 mg	race	ASIAN	7 (21.9)
30 mg: Placebo	race	ASIAN	5 (15.6)
Placebo: 30 mg	race	ASIAN	6 (18.8)
40 mg: Placebo	race	ASIAN	5 (16.1)
Placebo: 40 mg	race	ASIAN	5 (16.1)
All	race	ASIAN	44 (17.3)
10 mg: Placebo	race	WHITE N(%)	23 (74.2)
Placebo: 10 mg	race	WHITE	22 (68.8)
20 mg: Placebo	race	WHITE	21 (63.6)
Placebo: 20 mg	race	WHITE	21 (65.6)
30 mg: Placebo	race	WHITE	22 (68.8)
Placebo: 30 mg	race	WHITE	23 (71.9)
40 mg: Placebo	race	WHITE	23 (74.2)
Placebo: 40 mg	race	WHITE	23 (74.2)
All	race	WHITE	178 (70.1)
10 mg: Placebo	sex	FEMALE N(%)	14 (45.2)
Placebo: 10 mg	sex	FEMALE	21 (65.6)
20 mg: Placebo	sex	FEMALE	18 (54.5)
Placebo: 20 mg	sex	FEMALE	25 (78.1)
30 mg: Placebo	sex	FEMALE	15 (46.9)
Placebo: 30 mg	sex	FEMALE	21 (65.6)
40 mg: Placebo	sex	FEMALE	16 (51.6)
Placebo: 40 mg	sex	FEMALE	18 (58.1)
All	sex	FEMALE	148 (58.3)
10 mg: Placebo	sex	MALE N(%)	17 (54.8)
Placebo: 10 mg	sex	MALE	11 (34.4)
20 mg: Placebo	sex	MALE	15 (45.5)
Placebo: 20 mg	sex	MALE	7 (21.9)
30 mg: Placebo	sex	MALE	17 (53.1)
Placebo: 30 mg	sex	MALE	11 (34.4)
40 mg: Placebo	sex	MALE	15 (48.4)
Placebo: 40 mg	sex	MALE	13 (41.9)
All	sex	MALE	106 (41.7)

\*The other races represented in the study: African American and Multi-Racial are not shown in this table as the numbers were very small

There did not appear to be any significant baseline differences between assigned treatment sequences in terms of WASO or LPS.

**Table 18 Study 06: Baseline Disease Characteristics by Sequence**

Arm	Baseline Variable	N	Mean (Std. Dev.)
10 mg: Placebo	waso1	31	107.8 (53.3)
Placebo: 10 mg	waso1	32	97.9 (34.3)
20 mg: Placebo	waso1	33	95.7 (36.8)
Placebo: 20 mg	waso1	32	99.5 (43.5)
30 mg: Placebo	waso1	32	109.7 (50.9)
Placebo: 30 mg	waso1	32	104.8 (60.6)
40 mg: Placebo	waso1	31	93.7 (31.6)
Placebo: 40 mg	waso1	31	92.3 (41.7)
All	waso1	254	100.2 (44.9)
10 mg: Placebo	lps	31	65.8 (44.3)
Placebo: 10 mg	lps	32	74.9 (38.0)
20 mg: Placebo	lps	33	75.5 (34.3)
Placebo: 20 mg	lps	32	65.0 (33.6)
30 mg: Placebo	lps	32	71.9 (61.1)
Placebo: 30 mg	lps	32	59.0 (37.2)
40 mg: Placebo	lps	31	66.5 (37.8)
Placebo: 40 mg	lps	31	70.4 (36.6)
All	lps	254	68.7 (41.0)

### 3.1.3.4 Sponsor's Results

#### Changes to the Analysis Plan

The protocol stated that a mixed model which included terms for: baseline value, region (Japan, ex-Japan), treatment, sequence, period, time (as a categorical variable), treatment by- time and period-by-time interactions, would be used for the evaluation of the primary and secondary endpoints. A decision was made prior to unblinding the data for the first interim analysis to multiply all terms in the model, with the exception of the treatment and treatment-by-time terms, by an indicator variable for 2-period crossover study (e.g. the MK-4305 10 mg: placebo and placebo:MK-4305 10 mg sequences comprised the 2- period crossover study for the MK-4305 10 mg dose). This change to the model still allowed placebo information to be pooled across the four 2-period crossover studies while estimating all other effects separately for each 2-period study. This may be important since these other terms could potentially vary with dose.

#### Interim Analyses

An interim analysis for futility of the MK-4305 doses was conducted when approximately 50% of patients completed the study. The analysis included only patients from the US cohort; no patients of Japanese heritage were included. Only an unblinded statistician and a standing internal data monitoring committee (siDMC) were unblinded for this analysis. Guidelines which documented the procedures, methods, and criteria for actions were prespecified in the siDMC charter for this interim analysis.

A second interim analysis was planned and conducted when there was approximately 80% power for the primary efficacy hypothesis (i.e., when approximately 160 patients completed both

periods of the crossover study). The purpose of this analysis was to expedite planning of the Phase III program. Both efficacy and safety were evaluated and the analysis included patients from the US cohort only. Guidelines for the procedures, methods and criteria for action were also prespecified in the same siDMC charter used for the 50% interim analysis. At the conclusion of this interim analysis, the siDMC directed the release of group-level information to the Merck team responsible for planning Phase III to select doses and begin preparation of the Phase III program; the team was not given access to the individual patient data. No other individuals within or outside of Merck were informed of the results of this interim analysis until the last patient had completed the study. There was no plan to end the trial for superior efficacy on the basis of this analysis and the method of Haybittle-Peto was used to account for multiplicity related to this evaluation.

### Efficacy Summary for Periods 1 and 2 Combined

Based on the testing strategy:

1. All doses of MK-4305 (i.e., 80 mg, 40 mg, 20 mg and 10 mg) were more effective than placebo in improving insomnia as measured by the primary efficacy endpoint, sleep efficiency (SE), at Night 1 and at the end of Week 4. 2. All doses of MK-4305 were more effective than placebo in improving sleep maintenance as measured by the secondary efficacy endpoint, wakefulness after persistent sleep onset (WASO), at Night 1 and at the end of Week 4. 3. No doses of MK-4305 were more effective than placebo in improving sleep onset as measured by the secondary efficacy endpoint, latency to onset of persistent sleep (LPS), at Night 1 and at the end of Week 4, according to the multiplicity testing strategy. However, all doses of MK-4305 had numeric decreases in LPS and multiple MK-4305 doses had nominal p-values for comparisons versus placebo which were < 0.001: namely, 80 mg and 40 mg on Night 1 and 20 mg at Week 4.

**Table 19 Study 06: P-value Summary of Sponsor’s Primary and Key Secondary Analyses**

Comparison	Primary Endpoints		Secondary Endpoints			
	SE		WASO		LPS	
	Night 1	Week 4	Night 1	Week 4	Night 1	Week 4
MK-4305 80 mg vs. Placebo	<.001	<.001	<.001	<.001	<0.001	0.068
MK-4305 40 mg vs. Placebo	<.001	<.001	<.001	<.001	<0.001	<0.459
MK-4305 20 mg vs. Placebo	<.001	<.001	<.001	<.001	0.130	<0.001
MK-4305 10 mg vs. Placebo	0.002	0.003	<.001	0.001	0.577	0.644

**P-values in bold are significant according to the multiplicity testing strategy.**

Note: this table was copied from page 121 of sponsor’s study report

### SE on Night 1 (Periods 1 and 2 Combined)

At baseline, the mean for SE was between 65 and 67 for all treatments. The estimated differences in SE between Night 1 and baseline were: 10.9, 17.8, 17.4, 23.7 and 21.8, for placebo, MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. These changes from baseline correspond to percent increases in SE relative to baseline of 17% for placebo (= 100 x 10.9/65.9), and 27%, 26%, 37%, and 32% for MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. The differences in Least Squares (LS) means for SE between MK-4305 and placebo on Night 1 were

all significant (p-value < 0.002); the placebo-subtracted differences were 5.2, 7.6, 10.8 and 12.9 for 10 mg, 20 mg, 40 mg and 80 mg, respectively.

#### **SE at Week 4 (Periods 1 and 2 Combined)**

The estimated differences in SE between Week 4 and baseline were: 12.3, 18.7, 19.1, 20.4 and 19.8, for placebo, MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. These changes from baseline correspond to percent increases in SE relative to baseline of 19% for placebo (= 100 x 12.3/65.4), and 28%, 29%, 32%, and 30% for MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively.

The differences in LS means for SE between MK-4305 and placebo at Week 4 were all significant (p-value  $\leq$  0.003); the differences were 4.7, 10.4, 7.8 and 7.6 between 10 mg, 20 mg, 40 mg and 80 mg and placebo, respectively, indicating that patients had significantly greater improvements in SE at Week 4 while on MK-4305 than while on placebo.

#### **WASO**

The differences in LS means for WASO between MK-4305 and placebo were all significant (p-value < 0.001); the differences were -21.2, -24.7, -33.9 and -36.8 minutes between 10 mg, 20 mg, 40 mg, 80 mg and placebo, respectively. Thus, all doses of MK- 4305 were superior to placebo with respect to sleep maintenance as measured by WASO.

The differences in LS means for WASO between MK-4305 and placebo were all significant (p-value < 0.001); the placebo-subtracted differences were -21.4, -28.1, -33.2, and -28.9 minutes between 10 mg, 20 mg, 40 mg and 80 mg, respectively, indicating that patients had significantly greater improvement in sleep maintenance while on MK-4305 than while on placebo (see Table 20).

**Table 20 Study 06: Sponsor’s Analysis of WASO Based on Both Periods**

Treatment	N	Baseline Mean (SD)	Treatment Mean (SD)	Change from Baseline	
				Mean (SD)	95% CI
<b>Night 1</b>					
Placebo	249	100.7 (45.11)	70.6 (53.77)	-30.1 (58.25)	(-37.3, -22.8)
MR4305 10 mg	62	103.4 (44.70)	54.5 (37.86)	-48.9 (52.94)	(-62.3, -35.5)
MR4305 20 mg	61	97.7 (41.10)	46.6 (42.62)	-51.2 (38.42)	(-61.0, -41.3)
MR4305 40 mg	59	107.3 (57.58)	35.4 (20.51)	-71.8 (52.79)	(-85.6, -58.1)
MR4305 80 mg	61	93.9 (36.24)	27.3 (17.14)	-66.6 (36.04)	(-75.9, -57.4)
<b>Week 4</b>					
Placebo	232	101.7 (45.74)	73.7 (54.73)	-28.1 (62.81)	(-36.2, -19.9)
MR4305 10 mg	59	101.1 (41.40)	51.2 (31.91)	-49.9 (51.19)	(-63.2, -36.6)
MR4305 20 mg	57	97.1 (41.82)	47.0 (33.18)	-50.1 (37.53)	(-60.1, -40.2)
MR4305 40 mg	57	109.0 (57.84)	49.1 (40.43)	-59.9 (68.97)	(-78.2, -41.6)
MR4305 80 mg	55	94.5 (34.97)	40.5 (27.22)	-53.9 (43.15)	(-65.6, -42.3)
Pairwise Comparison		Differences in LS Mean (SE) <sup>†</sup>		95% CI	p-Value
<b>Night 1</b>					
MR4305 10 mg		-21.2 ( 6.27)		(-33.5, -8.8)	<.001
MR4305 20 mg		-24.7 ( 6.31)		(-37.1, -12.3)	<.001
MR4305 40 mg		-33.9 ( 6.33)		(-46.4, -21.5)	<.001
MR4305 80 mg		-36.8 ( 6.36)		(-49.4, -24.3)	<.001
<b>Week 4</b>					
MR4305 10 mg		-21.4 ( 6.46)		(-34.2, -8.7)	0.001
MR4305 20 mg		-28.1 ( 6.58)		(-41.0, -15.1)	<.001
MR4305 40 mg		-33.2 ( 6.61)		(-46.3, -20.2)	<.001
MR4305 80 mg		-28.9 ( 6.70)		(-42.1, -15.7)	<.001
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, sequence, period, time, treatment-by-time and period-by-time interactions; all terms are specific to each 2x2 crossover study (i.e., crossed with indicator variables for each 2x2 crossover study) with the exception of the treatment effect so as to pool placebo information across the 2x2 crossover studies. N = Number with measurement at time point.					

Note: This Table was copied from page 126 of sponsor’s study report

BEST AVAILABLE COPY

Table 21 Study 06 Sponsor's exploratory Analysis of WASO for first period only

Treatment	N	Baseline Mean (SD)	Treatment Mean (SD)	Change from Baseline	
				Mean (SD)	95% CI
<b>Night 1</b>					
Placebo	127	98.7 (43.76)	76.6 (39.07)	-22.1 (62.86)	(-33.1, -11.0)
MRE4305 10 mg	31	107.8 (53.28)	61.0 (40.80)	-46.9 (57.99)	(-68.1, -25.6)
MRE4305 20 mg	33	93.7 (36.80)	51.7 (30.57)	-44.0 (36.61)	(-57.0, -31.0)
MRE4305 40 mg	32	109.7 (50.93)	40.8 (23.54)	-69.0 (44.78)	(-85.1, -52.8)
MRE4305 80 mg	31	93.7 (31.60)	28.5 (18.13)	-65.2 (34.80)	(-78.0, -52.5)
<b>Week 4</b>					
Placebo	116	100.2 (46.91)	74.8 (54.53)	-25.4 (61.55)	(-36.7, -14.1)
MRE4305 10 mg	29	104.4 (47.91)	36.2 (38.21)	-68.2 (58.72)	(-70.5, -25.9)
MRE4305 20 mg	31	95.7 (37.86)	43.2 (30.06)	-52.5 (35.16)	(-65.4, -39.6)
MRE4305 40 mg	30	113.1 (50.76)	61.4 (48.41)	-51.7 (73.83)	(-80.0, -23.4)
MRE4305 80 mg	28	97.9 (30.29)	39.8 (21.71)	-58.1 (38.86)	(-73.2, -43.0)
Pairwise Comparison		Differences in LS Mean (SE) <sup>†</sup>		95% CI	p-Value
<b>Night 1</b>					
MRE4305 10 mg		-18.1 ( 9.36)		(-36.5, 0.4)	0.053
MRE4305 20 mg		-24.2 ( 9.12)		(-42.2, -6.2)	0.008
MRE4305 40 mg		-38.8 ( 9.25)		(-57.0, -20.5)	<.001
MRE4305 80 mg		-46.7 ( 9.35)		(-65.2, -28.3)	<.001
<b>Week 4</b>					
MRE4305 10 mg		-18.8 ( 9.46)		(-37.4, -0.2)	0.048
MRE4305 20 mg		-30.2 ( 9.21)		(-48.4, -12.0)	0.001
MRE4305 40 mg		-16.7 ( 9.36)		(-35.1, 1.8)	0.076
MRE4305 80 mg		-34.2 ( 9.58)		(-53.0, -15.3)	<.001
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, sequence, period, time, treatment-by-time and period-by-time interactions; all terms are specific to each 2x2 crossover study (i.e., crossed with indicator variables for each 2x2 crossover study) with the exception of the treatment effect so as to pool placebo information across the 2x2 crossover studies. N = Number with measurement at time point.					

Note: this table was copied from page 331 of sponsor's study report

*Reviewer's Comment:*

*There was no first period only analysis prespecified for WASO. For the first period only analysis the sponsor chose not to include the dose by covariate interactions that were prespecified for the analysis of both periods. WASO was significant at Day 1  $p=.036$  but strictly not at Day 28  $p=0.066$  if interactions like those used in the primary analysis of both periods were included in the first period only model. Note that the analysis of both periods (Table 20) is the preferred analysis by this reviewer since it was prespecified and there was no treatment group by period interaction apparent.*

## LPS

As shown in Table 22 the differences in LS means for LPS between MK-4305 and placebo at Week 4 were -2.3, -22.3, -3.8, and -9.5 minutes for 10 mg, 20 mg, 40 mg and 80 mg and placebo, respectively, but none was significant according to the multiplicity testing strategy. MK-4035 20 mg was nominally significant (p-value < 0.001).

**Table 22 Study 06: Sponsor's Analysis of Latency to Persistent Sleep Based on Both Periods**

Treatment	N	Baseline Mean (SD)	Treatment Mean (SD)	Change from Baseline	
				Mean (SD)	95% CI
<b>Night 1</b>					
Placebo	249	69.3 (41.13)	46.2 (30.32)	-23.0 (33.14)	(-29.9, -16.1)
MK4305 10 mg	62	69.9 (41.23)	33.5 (25.86)	-36.4 (46.39)	(-48.2, -24.6)
MK4305 20 mg	61	70.8 (33.12)	36.3 (31.62)	-34.5 (42.61)	(-45.4, -23.6)
MK4305 40 mg	39	66.8 (52.22)	23.7 (16.08)	-43.1 (50.80)	(-56.3, -29.8)
MK4305 80 mg	61	69.1 (36.80)	28.8 (22.31)	-40.3 (37.59)	(-50.0, -30.7)
<b>Week 4</b>					
Placebo	232	70.2 (41.51)	38.4 (44.84)	-31.8 (53.60)	(-39.0, -24.6)
MK4305 10 mg	39	69.7 (41.55)	28.0 (26.97)	-41.7 (45.41)	(-53.5, -29.8)
MK4305 20 mg	37	72.2 (33.62)	28.0 (18.21)	-44.2 (35.97)	(-53.7, -34.7)
MK4305 40 mg	37	66.2 (51.82)	28.0 (25.08)	-38.2 (51.12)	(-51.8, -24.6)
MK4305 80 mg	35	72.8 (36.43)	29.8 (29.44)	-43.0 (38.67)	(-53.4, -32.5)
Pairwise Comparison		Differences in LS Mean (SE) <sup>†</sup>		95% CI	p-Value
<b>Night 1</b>					
MK4305 10 mg		-3.4 ( 6.16)		(-15.6, 8.7)	0.577
MK4305 20 mg		-9.4 ( 6.17)		(-21.5, 2.6)	0.130
MK4305 40 mg		-23.1 ( 6.20)		(-35.3, -10.9)	<.001
MK4305 80 mg		-25.4 ( 6.23)		(-37.7, -13.1)	<.001
<b>Week 4</b>					
MK4305 10 mg		-2.3 ( 4.99)		(-12.2, 7.5)	0.644
MK4305 20 mg		-22.3 ( 5.08)		(-32.3, -12.3)	<.001
MK4305 40 mg		-3.8 ( 5.09)		(-13.8, 6.3)	0.439
MK4305 80 mg		-9.5 ( 5.17)		(-19.7, 0.7)	0.068
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, sequence, period, time, treatment-by-time and period-by-time interactions; all terms are specific to each 2x2 crossover study (i.e., crossed with indicator variables for each 2x2 crossover study) with the exception of the treatment effect so as to pool placebo information across the 2x2 crossover studies. N = Number with measurement at time point.					

BEST  
AVAILABLE  
COPY

Note: This table was copied from page 129 of the sponsor's study report

## Consideration of Carryover

Evidence of a potential carryover effect was observed for LPS in this 2-period crossover study. Patients who received placebo in Period 1 had further improvement in LPS when they received MK-4305 in Period 2; however, for patients who received MK-4305 in Period 1 improvement in LPS did not diminish in Period 2, even though patients received placebo in Period 2 (see Figure 11 on page 60). To further evaluate the efficacy of MK-4305 on LPS without the influence of carryover, an ad hoc analysis of LPS restricted to Period 1 was also performed.

### **LPS on Night 1 (Period 1 Only)**

The estimated differences in LPS between Night 1 in Period 1 and baseline were: -14.4, -32.2, -38.2, -49.2 and -36.0 minutes for placebo, MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. These changes from baseline correspond to percent decreases in LPS relative to baseline of 21% for placebo ( $= 100 \times 14.3/67.3$ ), and 49%, 51%, 68% and 54% for MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. The differences in LS means for LPS between MK-4305 10 mg, 20 mg, 40 mg, 80 mg and placebo in Period 1 were: -19.1, -17.4, -31.0, and -22.3 minutes, respectively. The difference between MK-4305 80 mg and placebo at Night 1 was significant (p-value = 0.007) according to the multiplicity testing strategy, indicating that MK-4305 80 mg significantly improved sleep onset on Night 1 in Period 1 as compared to placebo. There was also evidence of improvement of sleep onset with MK-4305 10 mg, 20 mg and 40 mg on Night 1 in Period 1, since the nominal p-values for comparison of these MK-4305 doses with placebo were all  $\leq 0.03$ .

### **LPS at Week 4 (Period 1 Only)**

The estimated differences in LPS between Week 4 in Period 1 and baseline were: -19.2, -37.1, -52.0, -38.1 and -40.9 minutes for placebo, MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively (Table 11-6, Figure 11-4). These changes from baseline correspond to percent decreases in LPS relative to baseline of 28% for placebo ( $= 100 \times 19.2/67.8$ ), and 58%, 67%, 54% and 58% for MK-4305 10 mg, 20 mg, 40 mg and 80 mg, respectively. The differences in LS means for LPS between MK-4305 10 mg, 20 mg, 80 mg and placebo in Period 1 were -20.2, -24.6, -15.7, and -19.6 minutes, respectively. Only the difference between MK-4305 80 mg and placebo at Week 4 in Period 1 was significant (p-value = 0.024) according to the multiplicity testing strategy. This indicates that MK-4305 80 mg significantly improved sleep onset at Week 4 in Period 1 as compared to placebo. There was also possible evidence of improvement of sleep onset with the lower doses of MK-4305 at Week 4 in Period 1 as well, since the nominal p-values were both  $\leq 0.019$  for comparisons of 10 mg and 20 mg with placebo and the nominal p-value for MK-4305 40 mg versus placebo was nearly significant (p = 0.063).

### **Efficacy Summary of LPS for Period 1**

Evidence of a potential carryover effect for LPS was observed in this 2-period crossover study. Patients who received placebo in Period 1 did demonstrate further improvement in LPS when they received MK-4305 in Period 2; however, for patients who received MK-4305 in Period 1, improvement in LPS did not appear to diminish in Period 2 even though they received placebo in Period 2. An ad hoc analysis of LPS for Period 1 data only (i.e., eliminating the potential influence of carryover effects) applying the prespecified multiplicity testing strategy showed that based on this alternative analysis (see Table 23):

1. MK-4305 80 mg significantly improved sleep onset as compared to placebo.
2. While not statistically significant according to the multiplicity testing strategy, all doses of MK-4305 had numeric decreases in LPS which were greater in magnitude than those observed for placebo. The nominal p-values were  $< 0.05$  for nearly all comparisons of MK-4305 versus placebo; the only exception was the comparison of MK-4305 40 mg versus placebo at Week 4, which was nearly nominally significant (p-value = 0.063).

**Table 23 Study 06: Sponsor’s Analysis of Latency to Persistent Sleep Based on First Period Only**

Treatment	N	Baseline Mean (SD)	Treatment Mean (SD)	Change from Baseline	
				Mean (SD)	95% CI
<b>Night 1</b>					
Placebo	127	67.3 (36.45)	53.0 (51.20)	-14.3 (57.02)	(-24.3, -4.3)
MK4305 10 mg	31	65.8 (44.29)	33.6 (26.85)	-32.2 (49.84)	(-50.5, -14.0)
MK4305 20 mg	33	75.5 (34.33)	37.3 (37.03)	-38.2 (45.93)	(-54.5, -21.9)
MK4305 40 mg	32	71.9 (61.05)	22.7 (14.65)	-49.2 (59.32)	(-70.6, -27.8)
MK4305 80 mg	31	66.5 (37.77)	30.5 (23.99)	-36.0 (40.67)	(-51.0, -21.1)
<b>Week 4</b>					
Placebo	116	67.8 (36.55)	48.6 (53.96)	-19.2 (58.31)	(-30.0, -8.5)
MK4305 10 mg	29	64.3 (44.58)	27.2 (27.58)	-37.1 (46.83)	(-54.9, -19.2)
MK4305 20 mg	31	77.8 (34.13)	25.8 (17.99)	-52.0 (37.24)	(-65.7, -38.4)
MK4305 40 mg	30	71.2 (61.05)	33.0 (29.90)	-38.1 (59.92)	(-60.5, -15.8)
MK4305 80 mg	28	70.8 (37.20)	29.9 (25.68)	-40.9 (37.63)	(-55.5, -26.3)
Pairwise Comparison		Differences in LS Mean (SE) <sup>†</sup>		95% CI	p-Value
<b>Night 1</b>					
MK4305 10 mg		-19.1 ( 8.14)		(-35.1, -3.0)	0.020
MK4305 20 mg		-17.4 ( 7.96)		(-33.1, -1.7)	0.030
MK4305 40 mg		-31.0 ( 8.04)		(-46.9, -15.2)	<0.001
MK4305 80 mg		-22.3 ( 8.14)		(-38.3, -6.2)	0.007
<b>Week 4</b>					
MK4305 10 mg		-20.2 ( 8.52)		(-37.0, -3.4)	0.019
MK4305 20 mg		-24.6 ( 8.31)		(-41.0, -8.3)	0.003
MK4305 40 mg		-15.7 ( 8.41)		(-32.3, 0.8)	0.063
MK4305 80 mg		-19.6 ( 8.62)		(-36.6, -2.6)	0.024
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, time, and treatment-by-time interactions.					
N = Number with measurement at time point.					

Note: This table was copied from page 132 of the sponsor’s study report

**Reviewer’s Comment:** For this analysis of LPS in the first period only, the sponsor removed the interactions between dose and time, dose and baseline score, and dose and country from the model for some unexplained reason. The results for the first period only appear somewhat less favorable based on the analyses including these interactions in the model as the sponsor prespecified and did for the analysis of both periods. This is particularly true for 10 mg on both days and 40 mg at Day 28 where the LPS results are no longer significant when the interactions are involved (10 mg: -11.8, p=0.256 at Day 1 and -6.1, p=0.554 at Day 28; 40 mg at Day 28: -4.7, p=0.65).

Table 24 illustrates some unexpected significant variability between placebo groups associated with different treatment sequences in terms of mean LPS in the first period.

**Table 24 Study 06: Summary Statistics for Latency to Persistent Sleep in First Period**

Treatment	N	Baseline Mean (SD)	Treatment Mean (SD)	Change from Baseline	
				Mean (SD)	(95% CI)
<b>Period 1</b>					
<b>Night 1</b>					
Placebo for sequences with MK4305 10 mg	32	74.9 (38.00)	46.8 (40.44)	-28.1 (54.24)	(-47.7, -8.5)
Placebo for sequences with MK4305 20 mg	32	65.0 (33.64)	58.7 (52.95)	-6.2 (59.98)	(-27.8, 15.4)
Placebo for sequences with MK4305 40 mg	32	59.0 (37.23)	53.4 (57.06)	-5.6 (61.91)	(-27.9, 16.7)
Placebo for sequences with MK4305 80 mg	31	70.4 (36.55)	52.8 (54.61)	-17.5 (50.74)	(-36.1, 1.1)
Placebos pooled over all sequences	127	67.3 (36.45)	53.0 (51.20)	-14.3 (57.02)	(-24.3, -4.3)
MK4305 10 mg	31	65.8 (44.29)	33.6 (26.85)	-32.2 (49.84)	(-50.5, -14.0)
MK4305 20 mg	33	75.5 (34.33)	37.3 (37.03)	-38.2 (45.93)	(-54.5, -21.9)
MK4305 40 mg	32	71.9 (61.05)	22.7 (14.65)	-49.2 (59.32)	(-70.6, -27.8)
MK4305 80 mg	31	66.5 (37.77)	30.5 (23.99)	-36.0 (40.67)	(-51.0, -21.1)
<b>Week 4</b>					
Placebo for sequences with MK4305 10 mg	31	73.9 (38.22)	35.5 (31.48)	-38.4 (49.23)	(-56.4, -20.3)
Placebo for sequences with MK4305 20 mg	28	65.2 (31.32)	74.5 (81.00)	9.3 (68.65)	(-17.3, 35.9)
Placebo for sequences with MK4305 40 mg	29	59.7 (38.60)	36.4 (42.42)	-23.3 (57.80)	(-45.3, -1.3)
Placebo for sequences with MK4305 80 mg	28	72.0 (37.44)	49.7 (42.70)	-22.4 (48.18)	(-41.1, -3.7)
Placebos pooled over all sequences	116	67.8 (36.55)	48.6 (53.96)	-19.2 (58.31)	(-30.0, -8.5)
MK4305 10 mg	29	64.3 (44.58)	27.2 (27.58)	-37.1 (46.83)	(-54.9, -19.2)
MK4305 20 mg	31	77.8 (34.13)	25.8 (17.99)	-52.0 (37.24)	(-65.7, -38.4)
MK4305 40 mg	30	71.2 (61.05)	33.0 (29.90)	-38.1 (59.92)	(-60.5, -15.8)
MK4305 80 mg	28	70.8 (37.20)	29.9 (25.68)	-40.9 (37.63)	(-55.5, -26.3)

Note: This table was copied from page 324 of the sponsor's study report

### Subjective Sleep Endpoints

Patients reported significantly better total sleep time, time to sleep onset, wake after sleep onset and subjective sleep quality for the two higher doses of MK-4305 (40 mg and 80 mg) than for placebo on most weeks.

*Reviewer's note: 10 and 20 mg doses did not show nominal significance compared to placebo on sTST or sTSO at any week except for 20 mg at week 2 for sTSO ( $p=0.0423$  however, 80 mg which had precedence in the hierarchy did not show nominal significance at that time  $p=0.0780$ ).*

Table 25 summarizes the sponsor's analysis of subjective TSO

**Table 25 Study 06: Sponsor's Analysis of Subjective TSO: Differences from Placebo**

Pairwise Comparison	Differences in LS Mean <sup>†</sup> (95% CI)	p-Value
<b>Week 1</b>		
MK4305 10 mg	-2.4 (-8.8, 4.0)	0.4592
MK4305 20 mg	-4.2 (-10.8, 2.3)	0.2057
MK4305 40 mg	-12.8 (-19.5, -6.0)	0.0002
MK4305 80 mg	-5.0 (-11.7, 1.7)	0.1396
<b>Week 2</b>		
MK4305 10 mg	1.9 (-4.5, 8.3)	0.5630
MK4305 20 mg	-6.8 (-13.3, -0.2)	0.0423
MK4305 40 mg	-12.2 (-19.0, -5.5)	0.0005
MK4305 80 mg	-6.0 (-12.7, 0.7)	0.0780
<b>Week 3</b>		
MK4305 10 mg	-0.7 (-6.8, 5.3)	0.8121
MK4305 20 mg	-5.8 (-12.0, 0.4)	0.0688
MK4305 40 mg	-13.2 (-19.6, -6.8)	<.0001
MK4305 80 mg	-11.1 (-17.5, -4.8)	0.0007
<b>Week 4</b>		
MK4305 10 mg	-3.0 (-9.3, 3.3)	0.3523
MK4305 20 mg	-4.3 (-10.8, 2.2)	0.1946
MK4305 40 mg	-17.4 (-24.1, -10.7)	<.0001
MK4305 80 mg	-7.7 (-14.3, -1.1)	0.0219
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, sequence, period, time, treatment-by-time and period-by-time interactions; all terms are specific to each 2x2 crossover study (i.e., crossed with indicator variables for each 2x2 crossover study) with the exception of the treatment effect so as to pool placebo information across the 2x2 crossover studies. N = Number with measurement at time point.		

BEST AVAILABLE COPY

Note: This table was copied from page 412 of the sponsor's study report

This reviewer also conducted first period only analyses for subjective Time to Sleep Onset (see Table 26) using the same model used by the sponsor for the objective endpoint first period only analyses.

**Table 26 Study 06: Analysis of sTSO in first period only**

<b>Dose Day</b>	<b>Diff. from Placebo</b>	<b>Std. Err.</b>	<b>p-value</b>
<b>10 1</b>	-3.7189	4.1225	0.3680
<b>10 28*</b>	-10.4748	5.2403	0.0468
<b>20 1</b>	-4.0995	4.0950	0.3179
<b>20 28</b>	-5.4338	5.2913	0.3056
<b>40 1</b>	-12.7047	4.1970	0.0028
<b>40 28</b>	-13.8431	5.3507	0.0103
<b>80 1</b>	-7.6403	4.2265	0.0720
<b>80 28</b>	-13.7256	5.4729	0.0129

\*This was not nominally significant  $p=0.14$ , when dose by covariate interactions as used in the prespecified analysis were added.

Table 27 summarizes the sponsor's analysis of subjective TST.

**Table 27 Study 06: Sponsor's Analysis of Subjective TST**

Pairwise Comparison	Differences in LS Mean <sup>†</sup> (95% CI)	p-Value
<b>Week 1</b>		
MK4305 10 mg	3.0 (-10.4, 16.4)	0.6607
MK4305 20 mg	-3.1 (-16.8, 10.6)	0.6570
MK4305 40 mg	22.8 (8.6, 36.9)	0.0018
MK4305 80 mg	20.8 (6.8, 34.8)	0.0039
<b>Week 2</b>		
MK4305 10 mg	0.6 (-12.2, 13.3)	0.9302
MK4305 20 mg	10.5 (-2.5, 23.6)	0.1137
MK4305 40 mg	18.6 (5.1, 32.1)	0.0072
MK4305 80 mg	22.5 (9.3, 35.8)	0.0010
<b>Week 3</b>		
MK4305 10 mg	1.2 (-11.6, 14.0)	0.8584
MK4305 20 mg	5.4 (-7.7, 18.5)	0.4171
MK4305 40 mg	21.4 (7.9, 34.9)	0.0020
MK4305 80 mg	22.5 (9.1, 35.8)	0.0011
<b>Week 4</b>		
MK4305 10 mg	5.5 (-6.3, 17.3)	0.3578
MK4305 20 mg	-1.8 (-13.9, 10.4)	0.7741
MK4305 40 mg	29.6 (17.1, 42.1)	<.0001
MK4305 80 mg	19.4 (7.1, 31.7)	0.0022
<sup>†</sup> Results based on a mixed effects model with terms for baseline value, region, treatment, sequence, period, time, treatment-by-time and period-by-time interactions; all terms are specific to each 2x2 crossover study (i.e., crossed with indicator variables for each 2x2 crossover study) with the exception of the treatment effect so as to pool placebo information across the 2x2 crossover studies. N = Number with measurement at time point.		

Note: This table was copied from page 410 of sponsor's study report

This reviewer also conducted first period only analyses for subjective Total Sleep Time (see Table 28).

**Table 28 Study 06: Analysis of sTST in first period only**

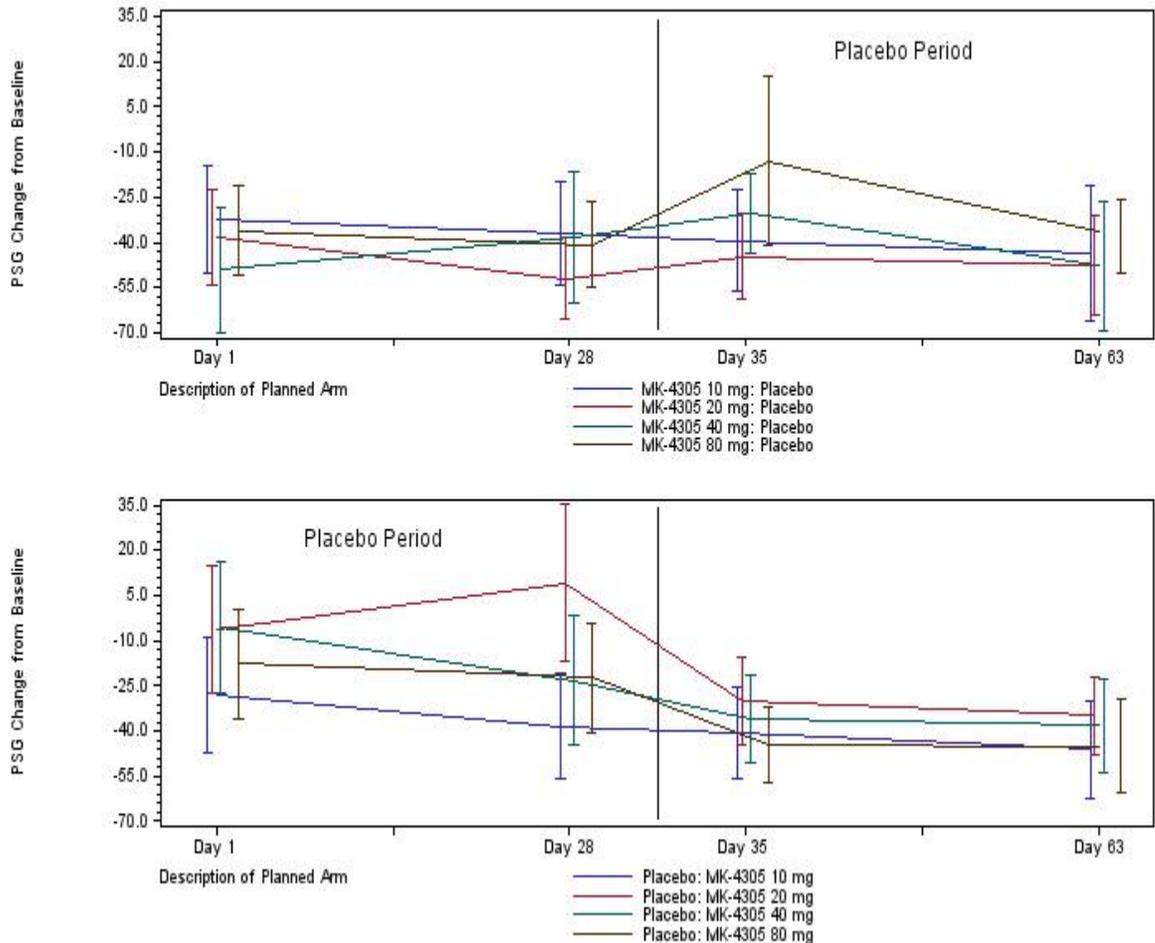
Dose Day	Diff. from Placebo	Std. Err.	p-value
<b>10 1</b>	5.5400	8.2282	0.5015
<b>10 28</b>	15.4204	9.3958	0.1022
<b>20 1</b>	5.1740	8.1668	0.5270
<b>20 28</b>	4.9265	9.4701	0.6034
<b>40 1</b>	19.8850	8.4068	0.0189
<b>40 28</b>	12.6206	9.6240	0.1911
<b>80 1</b>	18.9382	8.4577	0.0261
<b>80 28</b>	21.5596	9.8083	0.0290

### 3.1.3.5 Reviewer's Results

As discussed by the sponsor there is evidence of a carryover effect from period 1 into period 2 for LPS. The interaction effect between period and treatment had a p-value of 0.0118, suggesting that the LPS's of some treatments arms varied significantly across the two periods.

This reviewer found that based only on the 2 sequences involving 10 mg, the 10 mg effects at Day 1 and Day 28 on LPS were not nominally significant: -3.45 +/- 4.68 (S.E.), p=.4633 at Day 1 and -1.35 +/- 3.43(S.E.), p=.6945 at Day 28. Also, based only on these 2 sequences there was no evidence of a carryover effect on LPS.

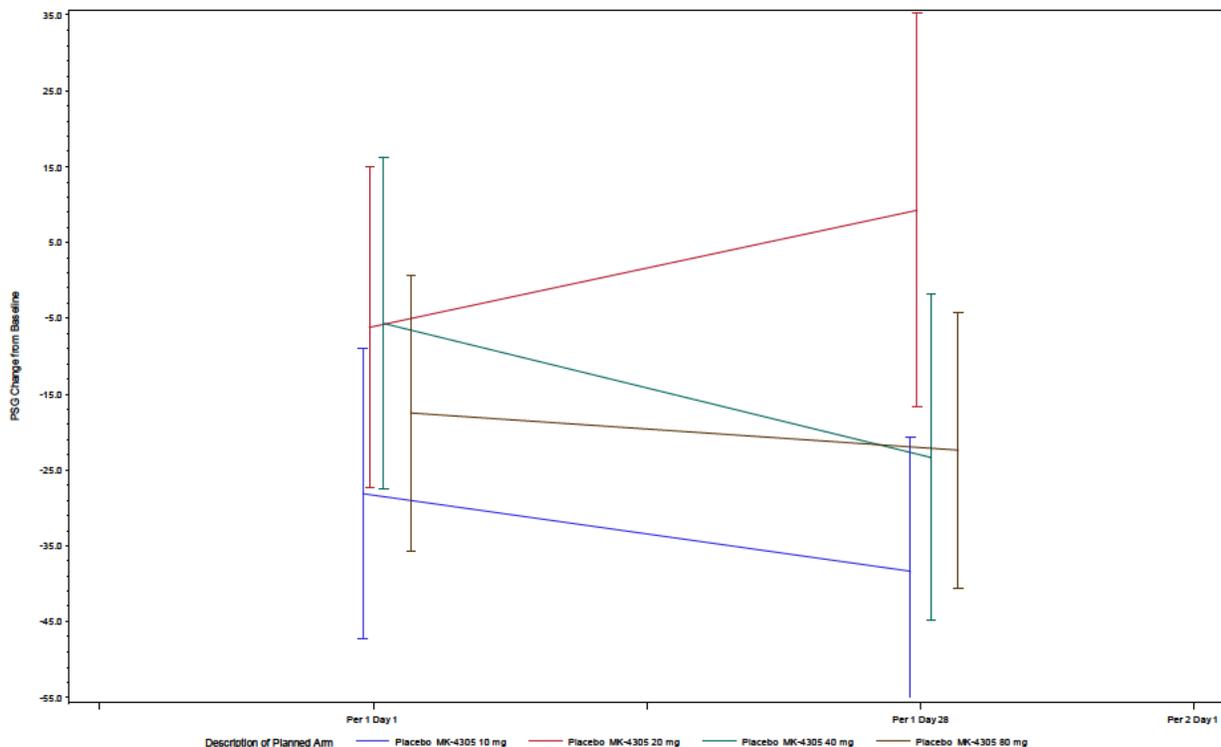
Figure 11 Study 06: Mean Change from Baseline in LPS by Sequence and Period



For LPS there was a nominally significant difference between the sequences assigned placebo in the first period (see the lower left side of Figure 11 or see Figure 12). There was no nominally significant difference between these groups in terms of mean baseline LPS. The post-baseline difference was most apparent at the day 28 timepoint, ( $p=0.0003$  for day 28 only and  $p=0.0033$  over both days). This may be due to chance alone but while there was no similar effect on WASO there was a similar trend in Sleep Efficiency but it was less significant ( $p=0.0665$  at day 28 and  $p=0.2788$  over both days).

The LPS effect of 10 mg on Day 1 was estimated to be  $-11.1 \pm 10.3$  (S.E.) minutes ( $p=0.2806$ ) based on substudy 1 only (10 mg/placebo and placebo/10 mg sequences) first period. The LPS effect of 10 mg on Day 1 estimated based on pooling the placebo from all substudies was  $-21.7 \pm 8.4$  (S.E.) ( $p=0.0107$ ). The effect of LPS of 10 mg on Day 28 was estimated as  $-5.3867 \pm 10.3$ ,  $p=0.6007$  based on substudy 1 only and was  $-20.6 \pm 8.2423$ ,  $p=0.0131$  based on pooled placebo. If the placebo corresponding to 20 mg was left out of the comparison the estimated effect of 10 mg at day 28 was  $-11.7$  ( $p=0.169$ ) and if placebo groups corresponding to both 10 mg and 20 mg were left out the estimated effect of 10 mg compared to the remaining placebos at day 28 was  $-14.9$  ( $p=0.101$ ). This highlights the sensitivity of the 10 mg effect to the placebo groups involved in the comparison.

**Figure 12 Study 06: Mean Change from Baseline in LPS during first period for the placebo first sequences**



BEST AVAILABLE COPY

A Wilcoxon rank sum test done as a post-hoc sensitivity analysis which may lessen the impact of any outliers gave a result of  $p=0.276$  for 10 mg vs. pooled placebo in the first period on Day 1 and  $p=0.075$  for Day 28.

Twenty (8%) of 254 randomized patients had missing data at Week 4 in the first period (10/P: N=2, P/10: 1, 20/P: 2, P/20: 4, 40/P:2, P/40:3, 80/P 3, P/80: 3) In the second period the numbers missing (11%) at Week 4 were: (10/P: N=2, P/10: 2, 20/P: 3, P/20: 6, 40/P:3, P/40:5, 80/P 3, P/80: 4).

This reviewer also performed sensitivity analyses for missing data using Multiple Imputation. The sponsor had prespecified such a model for their sensitivity analysis for missing data. The results are very similar to those for the observed data analyses, suggesting a limited impact of missing data on the results. The results of these analyses are shown for the 10 mg vs. placebo comparison in Table 29. I focus more on 10 mg here because 20 mg and 40 mg have more extensive data in the phase 3 studies but this is the only study involving 10 mg.

**Table 29 Study 06: Reviewer’s Multiple Imputation Sensitivity Analyses for Missing Data**

Efficacy Measure	Dose	Night	Difference from Placebo	Std. Error of Diff.	P-value
LPS	10	1	-19.04	8.15	0.0194 -
LPS	10	28	-22.04	8.80	0.0123 -
WASO	10	1	-18.01	9.37	0.0545
WASO	10	28	-19.71	9.85	0.0459

### 3.2 Evaluation of Safety

General Safety is not reviewed in this document. Please see the medical officer’s review for the evaluation of safety.

#### 3.2.1 Multiple Dose Studies to Evaluate Next Day Effects of MK-4305 (Suvorexant) on Driving Performance in Healthy Subjects

The effects of night time administration of suvorexant on next-morning driving performance were evaluated in two similarly designed highway driving studies in 24 healthy elderly ( $\geq 65$  years old, P039) and 28 healthy non-elderly (21-64 years old, P035) subjects. In both studies, driving performance was evaluated following single (on Day 2) and 8 consecutive nights (on Day 9) of suvorexant at low dose (15 mg in elderly and 20 mg in non-elderly) and high dose (30 mg in elderly and 40 mg in non-elderly) with the driving tests being conducted at ~9 hours post-dose. The primary endpoint was standard deviation of lane position (SDLP) which was used as a measure of driving performance. Zopiclone was used as active control in both studies. A published on-the-road highway driving study comparing the impairment caused by zopiclone in healthy subjects versus insomnia patients indicated that healthy subjects may be an appropriate population to evaluate effects of hypnotics on driving performance.

In both studies, zopiclone demonstrated assay sensitivity on both Day 2 and Day 9. In the elderly subjects, suvorexant (15 or 30 mg) did not result in impairment on next-day driving performance as assessed by mean SDLP (primary endpoint) and symmetry analysis of SDLP (secondary). In the non-elderly subjects, in the sponsor’s opinion there was no clinically meaningful impairment of next-day driving performance at either dose level (20 and 40 mg) since the 90% CI of SDLP was  $< 2.4$  cm (pre-specified clinical significance bound). However, except for 20 mg at Day 9 all were significantly worse than placebo. In addition, the symmetry analysis of SDLP revealed a statistically greater number of subjects with SDLP treatment difference of  $> 2.4$  cm (indicating impairment) than those with SDLP  $< -2.4$  cm on Day 2 for both suvorexant doses, and on Day 9 for 40 mg. The treatment effect was apparently less with the low dose than high dose of

suvorexant. The treatment difference on SDLP was smaller on Day 9 as compared to Day 2, suggesting to the sponsor the possibility of some tolerance effects after repeated dosing. Plasma concentrations at 11 hours post-dose were measured in both driving studies and the PK/PD relationship was explored for SDLP. There was an apparent dose response on SDLP. Study 35 had a few protocol deviations as follows. One subject, AN0005, was missing Day 9 following placebo treatment (period 3). This subject left the study site the evening before testing due to a panic attack. Subject AN0004 was missing driving data for the 20 mg suvorexant treatment (Period 1) due to a technical failure. Three subjects repeated one treatment period of the study. The repeated period data were used for each subject and the original period data were not used in the pharmacodynamics analysis due to an incomplete data set. The 3 subjects affected were [AN0011 (in Period 1, due to a technical failure of the car driving procedure); AN0016 (the subject was unable to return to the CRU on Day 8, Period 2 due to personal issues); and AN 0021 (in Period 4, due to a hardware issue occurring at the beginning of the driving assessment)]. Five subjects had prematurely stopped car driving test data. Of these 5 subjects, the car driving test data for 4 subjects (AN0002, AN0006, AN0007, and AN0021) were used in the analysis. AN0016 prematurely stopped the driving test during Day 2 Period 2; however, this subject repeated Period 2 as he was unable to attend Day 8. Only data from the repeated period was included in the analysis.

Table 30 summarizes the sponsor’s analyses of mean SDLP in the non-elderly study. These results were verified by this reviewer.

**Table 30 Study 35: Non-Elderly Driving Study Analysis of Mean SDLP**

Day	Treatment	N	LS Mean		Difference From Placebo	
			Mean	95% CI	Mean	90% CI
2	Placebo	28	15.53	(14.53,16.53)		
	Zopiclone 7.5 mg	28	17.66	(16.66,18.66)	2.14	(1.49,2.79)
	Suvorexant 20 mg	28	16.54	(15.54,17.54)	1.01	(0.36,1.66)
	Suvorexant 40 mg	28	17.19	(16.19,18.19)	1.66	(1.01,2.31)
9	Placebo	27	15.47	(14.46,16.47)		
	Zopiclone 7.5 mg	28	16.91	(15.91,17.91)	1.45	(0.79,2.10)
	Suvorexant 20 mg	28	15.94	(14.94,16.94)	0.48	(-0.18,1.13)
	Suvorexant 40 mg	28	16.77	(15.77,17.77)	1.31	(0.65,1.96)

Original Period 2, Day 2 data for AN0016 was not included in the analysis (the subject had a premature stop due to somnolence). The subject could not return for the Period 2, Day 8 visit and repeated this period. Only the repeat period is included in the analysis.

Note: This table was copied from page 60 of the sponsor’s study report

*Reviewer’s Comment:*

*At Day 2 both doses were significantly worse than placebo but the upper limits of the 90% C.I.s were below 2.4. At Day 9, the Mean SDLP was nominally significantly worse in the high dose than in the low dose as well as placebo (p=0.0006). On day 9 the 40 mg mean SDLP was*

*nominally significantly worse than that for 20 mg, 0.83 p=0.036. Also, the 40 mg 95% C.I.(.89,2.44) , just included the cutpoint of interest 2.40. On day 2, the 40 mg mean SDLP comparison to 20 mg had an estimated difference of 0.65 with a p-value of 0.101.*

Table 31 summarizes the symmetry analysis of subjects with differences from placebo > 2.40 or < -2.40 in the non-elderly study.

**Table 31 Study 35: Non-Elderly Driving Study Symmetry Analysis**

Treatment	Day=2			Day=9		
	Suvorexant 20 mg	Suvorexant 40 mg	Zopiclone 7.5 mg	Suvorexant 20 mg	Suvorexant 40 mg	Zopiclone 7.5 mg
n+1 <sup>†</sup>	6	10	14	2	6	8
n-1 <sup>‡</sup>	0	2	1	1	0	0
m = n+1 + n-1	6	12	15	3	6	8
Test Statistic <sup>§</sup>	2.45	2.31	3.36	0.58	2.45	2.83
Reject Null Hypothesis? <sup>  </sup>	Yes	Yes	Yes	No	Yes	Yes

<sup>†</sup>n+1 = number of subjects with treatment difference (from placebo) ≥ 2.4 cm.  
<sup>‡</sup>n-1 = number of subjects with treatment difference (from placebo) ≤ -2.4 cm.  
<sup>§</sup>Test Statistic = (n+1 - n-1)/sqrt(m).  
<sup>||</sup> Reject Null hypothesis if test statistic > 1.74, which is the critical value for general sign test with exact unconditional approach for sample size N=28 and N=27.  
Null Hypothesis: the percentage of subjects falling above the positive cut point and falling below the negative cut point is symmetric about zero.  
Original Period 2, Day 2 data for AN0016 was not included in the analysis (the subject had a premature stop due to somnolence). The subject could not return for the Period 2, Day 8 visit and repeated this period. Only the repeat period is included in the analysis.

Note: This table was copied from page 61 of the sponsor's study report

Table 32 summarizes the sponsor's analysis of mean SDLP for the elderly study, P039. The results were verified by this reviewer.

**Table 32 Study 39: Elderly Driving Study Analysis of Mean SDLP**

Day	Treatment	N	LS Mean		Difference from PBO	
			Mean	95% CI	Mean	90% CI
2	Placebo	24	16.67	(15.48,17.86)		
	Zopiclone 7.5 mg	24	18.56	(17.37,19.75)	1.89	(1.22,2.55)
	MK-4305 15 mg	24	16.24	(15.05,17.43)	-0.43	(-1.10,0.23)
	MK-4305 30 mg	24	17.04	(15.85,18.23)	0.37	(-0.30,1.03)
9	Placebo	24	15.41	(14.22,16.60)		
	Zopiclone 7.5 mg	24	16.58	(15.39,17.78)	1.17	(0.51,1.84)
	MK-4305 15 mg	24	15.50	(14.31,16.69)	0.09	(-0.58,0.76)
	MK-4305 30 mg	24	16.01	(14.82,17.20)	0.60	(-0.06,1.27)

Note: This table was copied from page 59 of the sponsor's study report

An exploratory comparison of 40 vs. 20mg on day 2 gave an estimate of 0.7992 (0.4027 S.E.), p= 0.0489. On day 9 the estimated difference between 40 mg and 20 mg was 0.51, p=0.205.

Table 33 summarizes the symmetry analysis of subjects with differences from placebo > 2.40 or < -2.40 in the elderly driving study.

**Table 33 Study 39: Elderly Driving Study Symmetry Analysis**

Treatment	Day=2			Day=9		
	MK-4305 15 mg	MK-4305 30 mg	Zopiclone 7.5 mg	MK-4305 15 mg	MK-4305 30 mg	Zopiclone 7.5 mg
n+1 †	0	3	8	0	5	6
n-1 ‡	3	1	0	0	1	1
m = n+1 + n-1	3	4	8	0	6	7
Test Statistic §	-1.73	1	2.83	N.D.	1.63	1.89
Reject Null Hypothesis?	No	No	Yes	No	No	Yes

† n+1 = number of subjects with treatment difference from placebo  $\geq 2.4$  cm.  
‡ n-1 = number of subjects with treatment difference from placebo  $\leq -2.4$  cm.  
§ Test Statistic =  $(n+1 - n-1)/\sqrt{m}$ .  
|| Reject Null hypothesis if test statistic > 1.74, which is the critical value for general sign test with exact unconditional approach for sample size N=24.  
Null Hypothesis: the percentage of subjects falling above the positive cut point and falling below the negative cut point is symmetric about zero.

Note: This table was copied from page 59 of the sponsor's study report

## 4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

### 4.1 Gender, Race, Age, and Geographic Region

#### 4.1.1 Gender

In study P028 North American sites accounted for 34% of all randomized patients, 42% were Elderly, 65% White and 26% Asian, 63% Female.

In P029 67% were Female, 80% were White, 8% Asian, and 41% were Elderly. Note that the p-values presented in section 4.1 and its subsections should be considered exploratory since they are not adjusted for the multiplicity of tests and in addition the study was not powered for assessing effects in these subgroups. Pooled study results are only considered if the contributing studies are positive on their own. In such cases these results are provided, assuming consistency across studies, to get a more precise estimate of the effect in subgroups. The LPS results for high dose should be considered exploratory for Month 3 since study 29 was not positive in terms of LPS at month 3. Also, low dose results should be considered exploratory since the studies allocated all of the alpha (type I error) to testing the high dose.

Study 28 and 29 were pooled for gender subgroup analyses since they were of very similar design. There was no compelling evidence of differential efficacy for Males and Females in terms of LPS, WASO, sTST, or sTSO (see Table 34, Table 35, Table 36, and Table 37).

**Table 34** Change from Baseline in LPS by Gender in P28/P29 Pooled

<b>Gender</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p- value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p- value</b>
Female	90	-5.2814	3.0223	0.0808	-7.6442	2.6106	0.0035
Female	30	-9.4241	2.8334	0.0009	-12.6431	2.4187	<.0001
Female	1	-13.2441	3.0688	<.0001	-17.5972	2.6243	<.0001
Male	90	-3.4120	4.1610	0.4124	-4.4094	3.5207	0.2106
Male	30	-8.6088	3.8437	0.0253	-9.6025	3.2989	0.0037
Male	1	-7.6409	4.1203	0.0639	-12.6783	3.5571	0.0004
Male vs. Female	90	1.8694	5.1410	0.7162	3.2348	4.3869	0.4610
Male vs. Female	30	0.8153	4.7733	0.8644	3.0406	4.0949	0.4579

**Table 35 Change from Baseline in WASO by Gender in P28/P29 Pooled**

<b>Gender</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
Female	90	-17.6079	5.2170	0.0008	-24.5786	4.8225	<.0001
Female	30	-26.0302	5.0257	<.0001	-26.0964	4.6486	<.0001
Female	1	-31.5294	4.3235	<.0001	-33.5833	3.9419	<.0001
Male	90	-14.6121	7.1993	0.0428	-20.1741	6.2150	0.0012
Male	30	-26.9247	6.9177	0.0001	-26.2577	5.9591	<.0001
Male	1	-34.5159	5.8526	<.0001	-46.2136	5.0503	<.0001
Male vs. Female	90	2.9957	8.8987	0.7365	4.4045	7.8763	0.5762
Male vs. Female	30	-0.8944	8.5561	0.9168	-0.1612	7.5662	0.9830

**Table 36 Change from Baseline in sTSO by Gender in P28/P29 Pooled**

<b>Gender</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
Female	90	-4.6446	2.7227	0.0882	-11.5033	2.4059	<.0001
Female	30	-3.8226	2.7690	0.1676	-11.3188	2.4485	<.0001
Female	1	-6.6865	2.2562	0.0031	-9.4449	1.9942	<.0001
Male	90	-8.3626	3.6853	0.0234	-9.5179	3.2091	0.0031
Male	30	-8.7585	3.7227	0.0187	-7.8697	3.2663	0.0161
Male	1	-4.9561	3.0244	0.1014	-9.0074	2.6601	0.0007
Male vs. Female	90	-3.7180	4.5824	0.4173	1.9854	4.0140	0.6209
Male vs. Female	30	-4.9360	4.6403	0.2876	3.4491	4.0852	0.3986

**Table 37 Change from Baseline in sTST by Gender in P28/P29 Pooled**

<b>Gender</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
Female	90	16.3379	4.3180	0.0002	26.0808	3.8157	<.0001
Female	30	19.0584	3.9626	<.0001	26.0755	3.5029	<.0001
Female	1	16.1804	3.2194	<.0001	25.0402	2.8458	<.0001
Male	90	15.0096	5.8447	0.0103	15.1642	5.0897	0.0029
Male	30	16.7534	5.3248	0.0017	16.6413	4.6734	0.0004
Male	1	12.4592	4.3166	0.0039	21.3010	3.7968	<.0001
Male vs. Female	90	-1.3283	7.2669	0.8550	-10.9166	6.3654	0.0865
Male vs. Female	30	-2.3050	6.6379	0.7284	-9.4342	5.8449	0.1067

In the crossover study (P06) 58% of the 254 randomized patients were Female. Although the difference between Males and Female estimated 10 mg effects on WASO at Day 1 of the first period is noticeably large it is not significant considering the standard error of the estimates. Overall, there was no compelling evidence of differential estimated effects between genders for WASO or LPS in study 06, first period.

**Table 38 Study 06 First Period Only Analyses of Change from Baseline in WASO and LPS by Gender**

<b>Endpoint</b>	<b>Gender</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>DF</b>	<b>t Value</b>	<b>Pr &gt;  t </b>
<b>WASO</b>	<b>Female</b>	<b>1</b>	-25.9918	13.5315	251	-1.92	0.0559
	<b>Female</b>	<b>28</b>	-20.1399	14.1788	251	-1.42	0.1567
	<b>Male</b>	<b>1</b>	-4.8758	13.4528	251	-0.36	0.7173
	<b>Male</b>	<b>28</b>	-16.6700	13.3522	251	-1.25	0.2130
<b>LPS</b>	<b>Female</b>	<b>1</b>	-19.1324	11.8231	251	-1.62	0.1069
	<b>Female</b>	<b>28</b>	-27.4794	12.7650	251	-2.15	0.0323
	<b>Male</b>	<b>1</b>	-17.0123	11.7838	251	-1.44	0.1501
	<b>Male</b>	<b>28</b>	-13.0058	12.0424	251	-1.08	0.2812

### 4.1.2 Race

In study P028 overall proportions of randomized patients a given race were: 65% White, 26% Asian, 6% Black, 3% Multi-Racial, and .1% Pacific Islander. Although a considerable overall proportion of patients were Asian they were almost exclusively enrolled in the Q only cohort so that while for subjective endpoints the proportion Asian was 26%, for PSG endpoints the proportion Asian was only 2%.

There was no suggestion of an interaction in study 28 for sTSO, while for sTST the ‘Other’ category was numerically in the wrong direction so there was a suggestion of an interaction at Month 1. However, at month 3 as well as Night 1 the ‘Other’ category was numerically better than placebo thereby diminishing the credibility of the interaction at 1 Month.

**Table 39 Study 28 Differences in Change from Baseline in sTSO High Dose vs. Placebo at Month 1 by Race**

<b>Subgroup or Contrast</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Asian</b>	-6.9274	4.9357	0.1608
<b>Other</b>	-4.0025	8.1062	0.6216
<b>White</b>	-8.1375	3.1227	0.0093
<b>Asian vs. White</b>	1.2102	5.8453	0.8360

There was no compelling evidence of a difference in effect of the high dose at 1 Month on sTSO between Whites and Asians (p=0.836) or Whites and ‘Others’ (p=0.886).

Table 40 shows Race subgroup analyses for subjective Total Sleep Time at 1 Month for the high dose. There was a suggestion of differential efficacy between Whites and Asians (p=0.0277) and Whites and ‘Others’ (p=.0162) at Month 1 for the high dose in the subjective Total Sleep Time. However, the estimated effect was still numerically in the right direction for Asians, the ‘Other’ subgroup was small, and this pattern was not exhibited on the other efficacy measures. Therefore, this pattern may be due to chance.

**Table 40 Study 28: Differences in Change from Baseline in sTST at Month 1 High Dose vs. Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Asian</b>	7.7452	7.5200	0.3033
<b>Other</b>	-2.2489	12.3660	0.8557
<b>White</b>	27.3731	4.7610	<.0001
<b>Asian vs. White</b>	19.6279	8.9013	0.0277

Table 41 shows Race subgroup analyses for objective LPS at Month 1 and Month 3 for the high dose. Note that the Questionnaire only (Q) cohort in study 28 consisted of only Japanese patients which reduces the PSG Asian cohort significantly enough to make it more appropriate to combine Asians with Others for race subgroup analyses involving PSG endpoints (only 2% were Asian in the PQ cohort).

**Table 41 Study 28: Differences in Change from Baseline in LPS for High Dose vs. Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	-3.8785	6.7906	0.5681
<b>Other</b>	30	8.3294	6.7368	0.2167
<b>White</b>	90	-11.0910	2.8294	<.0001
<b>White</b>	30	-14.6979	2.7945	<.0001
<b>Other vs. White</b>	90	-7.2124	7.3520	0.3269
<b>Other vs. White</b>	30	-23.0274	7.2873	0.0016

Table 42 shows Race subgroup analyses for objective WASO at Month 1 and Month 3 for the high dose. Effects were bigger in the ‘Other’ subgroup but were in the right direction and nominally significant for both subgroups.

**Table 42 Study 28: Differences in Change from Baseline in WASO for High Dose vs. Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	-41.4422	9.8198	<.0001
<b>Other</b>	30	-41.9337	9.4761	<.0001
<b>White</b>	90	-19.3328	4.0977	<.0001
<b>White</b>	30	-23.4159	3.9377	<.0001
<b>Other vs. White</b>	90	22.1094	10.6353	0.0380
<b>Other vs. White</b>	30	18.5178	10.2545	0.0714

In study P029 proportions of randomized patients a given race were: 80% white, 8% Asian, 5% Black, 7% Multi-Racial.

Table 43 shows subgroup estimates of the high dose effect on subjective TSO for race at Month 1 and Month 3. Both subgroups (White and Other) had high dose vs. placebo differences in the right direction and there was no compelling evidence of differential effect by subgroup (e.g., sTSOm interaction p=.51 at Day 30).

**Table 43 Study 29: sTSO Differences for High Dose vs Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	-14.6758	7.0811	0.0385
<b>Other</b>	30	-8.4170	6.8416	0.2189
<b>White</b>	90	-12.2893	3.4659	0.0004
<b>White</b>	30	-13.4379	3.3910	<.0001
<b>Other vs. White</b>	90	2.3865	7.8819	0.7621
<b>Other vs. White</b>	30	-5.0209	7.6332	0.5108

Table 44 shows subgroup estimates of the high dose effect on subjective TSO for race at Month 1 and Month 3.

**Table 44 Study 29: sTST Differences for High Dose vs Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	15.5220	10.4627	0.1383
<b>Other</b>	30	25.9098	9.1709	0.0048
<b>White</b>	90	26.9093	5.1278	<.0001
<b>White</b>	30	26.1581	4.5475	<.0001
<b>Other vs. White</b>	90	11.3874	11.6498	0.3286
<b>Other vs. White</b>	30	0.2482	10.2333	0.9807

Table 45 shows subgroup estimates of the high dose effect on objective WASO for race at Month 1 and Month 3.

**Table 45 Study 29: WASO Differences for High Dose vs Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	-18.4149	12.2228	0.1324
<b>Other</b>	30	-28.9747	11.6989	0.0135
<b>White</b>	90	-30.5603	3.9081	<.0001
<b>White</b>	30	-29.4636	3.8335	<.0001
<b>Other vs. White</b>	90	-12.1454	12.8323	0.3443
<b>Other vs. White</b>	30	-0.4889	12.3113	0.9683

Table 46 shows subgroup estimates of the high dose effect on objective LPS for race at Month 1 and Month 3.

**Table 46 Study 29: LPS Differences for High Dose vs. Placebo by Race**

<b>Subgroup or Contrast</b>	<b>Day</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
<b>Other</b>	90	10.0428	10.5963	0.3436
<b>Other</b>	30	-11.3683	9.3443	0.2242
<b>White</b>	90	-5.0456	3.4344	0.1423
<b>White</b>	30	-12.1714	3.0798	<.0001
<b>Other vs. White</b>	90	-15.0885	11.1401	0.1761
<b>Other vs. White</b>	30	-0.8032	9.8398	0.9350

In study P06 70% of randomized patients were White. Because of the small overall size of the study and the crossover design not much can be said about race subgroup effects from this study. Race was also mostly confounded with Country in this study (see regional estimates at the end of section 4.1.4).

### 4.1.3 Age

In studies 28 and 29 the mean age was 55 to 56 and about 58 to 60% were non-elderly.

Table 47 shows the results for mean change from baseline in LPS by Age group (Age<65 and Age ≥ 65). There was some suggestion of lower effects on LPS for Elderly at 1 month.

This was true both in the pooled analysis as well as for study 29 analyzed separately. For the study 29 only analysis the LPS month 1 apparent interaction between age groups for high dose alone vs. placebo had a p-value of 0.0071 or across both hi and low dose comparisons (p=0.0257). However, this may be a quantitative interaction though since the effect in the elderly subgroup was still numerically in the right direction (-2.99, p=.5031).

**Table 47 Change from Baseline in LPS by Age Group in P28/P29 Pooled**

<b>Age Group</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
Elderly	90	-6.3985	3.6977	0.0838	-7.8343	3.1984	0.0144
Elderly	30	-5.2240	3.4796	0.1335	-6.9298	2.9737	0.0199
Elderly	1	-10.1704	3.7511	0.0068	-17.6958	3.2263	<.0001
Non-Eld.	90	-3.2936	3.2468	0.3106	-5.4343	2.7649	0.0496
Non-Eld.	30	-12.1545	3.0090	<.0001	-15.1004	2.5723	<.0001
Non-Eld.	1	-12.0673	3.2611	0.0002	-14.5087	2.7907	<.0001
Eld. Vs. Non-Eld.	90	3.1050	4.9189	0.5280	2.4001	4.2296	0.5705
Eld. Vs. Non-Eld.	30	-6.9305	4.5985	0.1320	-8.1705	3.9337	0.0380

Table 48, Table 49, and Table 50 show the results by Age group for WASO, sTST, and sTSO, respectively. There was no compelling evidence of differential effects by Age group on these endpoints.

**Table 48 Change from Baseline in WASO by Age Group in P28/P29 Pooled Studies**

<b>Age Group</b>	<b>Day</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
Elderly	90	-14.9477	6.4178	0.0201	-16.8556	5.8105	0.0038
Elderly	30	-26.6104	6.2086	<.0001	-26.3445	5.5506	<.0001
Elderly	1	-39.7014	5.3117	<.0001	-45.4524	4.7546	<.0001
Non-Eld.	90	-17.9340	5.5967	0.0014	-27.4945	5.0231	<.0001
Non-Eld.	30	-26.0884	5.3627	<.0001	-26.3299	4.8434	<.0001
Non-Eld.	1	-27.1623	4.5850	<.0001	-33.1765	4.0860	<.0001
Eld. Vs. Non-Eld.	90	-2.9862	8.5167	0.7260	-10.6389	7.6854	0.1667
Eld. Vs. Non-Eld.	30	0.5221	8.2027	0.9493	0.0146	7.3693	0.9984

**Table 49 Change from Baseline in sTST by Age Group in P28/P29 Pooled**

Age Group	Day	Lo Diff. Est.	Lo Std. Err.	Lo p-value	Hi Diff. Est.	Hi Std. Err.	Hi p-value
Elderly	90	18.9499	5.3771	0.0004	20.6024	4.7279	<.0001
Elderly	30	15.5224	4.9435	0.0017	20.8347	4.3480	<.0001
Elderly	1	16.7522	4.0110	<.0001	25.1524	3.5307	<.0001
Non-Eld.	90	13.4874	4.5557	0.0031	23.2304	3.9986	<.0001
Non-Eld.	30	20.1079	4.1599	<.0001	23.9585	3.6685	<.0001
Non-Eld.	1	13.4727	3.3777	<.0001	22.6294	2.9783	<.0001
Eld. Vs. Non-Eld.	90	-5.4625	7.0395	0.4379	2.6280	6.1943	0.6714
Eld. Vs. Non-Eld.	30	4.5855	6.4523	0.4774	3.1238	5.6913	0.5832

**Table 50 Change from Baseline in sTSO by Age Group in P28/P29 Pooled**

Age Group	Day	Lo Diff. Est.	Lo Std. Err.	Lo p-value	Hi Diff. Est.	Hi Std. Err.	Hi p-value
Elderly	90	-6.8121	3.3914	0.0447	-9.2371	2.9814	0.0020
Elderly	30	-3.8744	3.4583	0.2627	-8.7419	3.0409	0.0041
Elderly	1	-6.7878	2.8099	0.0158	-9.6849	2.4733	<.0001
Non-Eld.	90	-5.5008	2.8735	0.0558	-11.9246	2.5215	<.0001
Non-Eld.	30	-6.9660	2.9091	0.0167	-11.0524	2.5664	<.0001
Non-Eld.	1	-5.7715	2.3659	0.0148	-9.0448	2.0860	<.0001
Eld. Vs. Non-Eld.	90	1.3113	4.4389	0.7677	-2.6875	3.9063	0.4916
Eld. Vs. Non-Eld.	30	-3.0915	4.5133	0.4934	-2.3106	3.9807	0.5617

Note: using studyid\*cohort interaction

Study 06 was completely non-elderly so the above subgroup comparisons were not possible for study 06.

#### 4.1.4 Region

In this section regional subgroup effects (North America vs. non-North America) are investigated. These subgroup estimates were determined by augmenting the primary analysis model with interactions between region and visit, region and treatment group, as well as region and the interaction of visit and treatment group.

Table 51 Study 29: Regional Subgroup Analyses for LPS

Region	Time	Lo Diff. Est.	Lo Std. Err.	Lo p-value	Hi Diff. Est.	Hi Std. Err.	Hi p-value
North Am	Month 1	-2.8517	4.7425	0.5478	-7.6357	3.7187	0.0404
Non-North Am	Month 1	-14.6974	12.4632	0.2387	-24.6394	9.3317	0.0085
North Am vs. non-North Am	Month 1	-11.8458	13.3330	0.3746	-17.0037	10.0471	0.0910
North Am	Month 3	1.1308	5.2819	0.8306	-3.6565	4.1703	0.3809
Non-North Am	Month 3	7.7795	15.4501	0.6148	-0.5123	10.4061	0.9608
North Am vs. non-North Am	Month 3	-6.6487	16.3252	0.6839	-3.1442	11.2128	0.7793

Table 52 Study 29 Regional Subgroup Analyses for WASO

Region	Time	Lo Diff. Est.	Lo Std. Err.	Lo p-value	Hi Diff. Est.	Hi Std. Err.	Hi p-value
North Am	Month 1	-21.6962	5.8802	0.0002	-34.7564	4.5887	<.0001
Non-North Am	Month 1	-19.0745	15.4411	0.2171	-1.2321	11.7438	0.9165
North Am vs. non-North Am	Month 1	2.6217	16.5212	0.8740	33.5243	12.6063	0.0080
North Am	Month 3	-33.0982	6.0060	<.0001	-32.9251	4.7571	<.0001
Non-North Am	Month 3	-28.6841	17.2760	0.0973	-10.3186	11.7657	0.3808
North Am vs. non-North Am	Month 3	-4.4141	18.2876	0.8093	-22.6065	12.6892	0.0753

**Table 53 Study 28 Regional Subgroup Analyses for LPS**

<b>Label</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	<b>Month 1</b>	-8.2783	4.2446	0.0515	-12.2996	3.8838	0.0016
<b>Non-North Am</b>	<b>Month 1</b>	-5.8140	7.2390	0.4221	-1.5975	5.6410	0.7771
<b>North Am vs. non-North Am</b>	<b>Month 1</b>	2.4643	8.3915	0.7691	10.7022	6.8454	0.1184
<b>North Am</b>	<b>Month 3</b>	-10.8238	4.2780	0.0116	-10.9020	3.9093	0.0054
<b>Non-North Am</b>	<b>Month 3</b>	-1.6839	7.2920	0.8174	-4.1835	5.6430	0.4587
<b>North Am vs. non-North Am</b>	<b>Month 3</b>	9.1399	8.4544	0.2800	6.7185	6.8619	0.3279

**Table 54 Study 28 Regional Subgroup Analyses for WASO**

<b>Region</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	<b>Month 1</b>	-37.7779	5.9557	<.0001	-35.4144	5.4580	<.0001
<b>Non-North Am</b>	<b>Month 1</b>	-19.6475	10.1875	0.0542	-24.3390	7.9152	0.0022
<b>North Am vs. non-North Am</b>	<b>Month 1</b>	18.1304	11.8061	0.1251	11.0754	9.6196	0.2500
<b>North Am</b>	<b>Month 3</b>	-22.0289	6.1558	0.0004	-28.1706	5.6365	<.0001
<b>Non-North Am</b>	<b>Month 3</b>	-20.7892	10.5205	0.0486	-26.0219	8.1967	0.0016
<b>North Am vs. non-North Am</b>	<b>Month 3</b>	1.2397	12.1972	0.9191	2.1487	9.9530	0.8291

Subjective TST (sTST) results for High Dose minus placebo at 1 month for North America and non-North America were as follows.

**Table 55 Study 28 Regional Subgroup Analyses for sTST**

<b>Region</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	Month 1	22.5926	7.2159	0.0018	23.6983	6.6089	0.0004
<b>Non-North Am</b>	Month 1	18.9616	9.7241	0.0515	6.3231	7.5127	0.4002
<b>North Am vs. non-North Am</b>	Month 1	-3.6310	12.1092	0.7644	-17.3752	10.0065	0.0828
<b>North Am</b>	Month 3	17.0706	7.4735	0.0226	26.4447	6.8328	0.0001
<b>Non-North Am</b>	Month 3	10.1800	10.0306	0.3104	6.9730	7.7541	0.3687
<b>North Am vs. non-North Am</b>	Month 3	-6.8906	12.5087	0.5819	-19.4717	10.3357	0.0599

**Table 56 Study 28 Regional Subgroup Analyses for sTSO**

<b>Region</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	Month 1	8.7769	4.7820	0.0667	-4.7262	4.3771	0.2805
<b>Non-North Am</b>	Month 1	-4.7922	6.4404	0.4570	-8.2226	4.9773	0.0988
<b>North Am vs. non-North Am</b>	Month 1	3.9847	8.0207	0.6194	-3.4964	6.6288	0.5980
<b>North Am</b>	Month 3	-9.0915	3.7102	0.0144	-12.1773	3.3904	0.0003
<b>Non-North Am</b>	Month 3	-3.1593	4.9681	0.5250	-6.7427	3.8639	0.0812
<b>North Am vs. non-North Am</b>	Month 3	5.9322	6.1992	0.3388	5.4346	5.1416	0.2907

Technical Note: Arh(1) covariance structure used because unstructured failed to converge

In study P029 about 48% of all randomized patients were associated with North American sites. Regional treatment effects for the key efficacy variables are provided in the following tables.

**Table 57 Study 29 Regional Subgroup Analyses for sTST**

<b>Region</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	Month 1	18.4111	6.9755	0.0084	20.8290	5.7079	0.0003
<b>Non-North Am</b>	Month 1	26.9385	7.3475	0.0003	29.2487	6.9217	<.0001
<b>North Am vs. non-North Am</b>	Month 1	8.5275	10.1176	0.3995	8.4198	8.9711	0.3482
<b>North Am</b>	Month 3	17.1261	7.8194	0.0288	18.3977	6.3935	0.0041
<b>Non-North Am</b>	Month 3	24.3438	8.3824	0.0038	28.0552	7.8412	0.0004
<b>North Am vs. non-North Am</b>	Month 3	7.2177	11.4511	0.5286	9.6575	10.1173	0.3400

**Table 58 Study 29 Regional Subgroup Analyses for sTSO**

<b>Region</b>	<b>Time</b>	<b>Lo Diff. Est.</b>	<b>Lo Std. Err.</b>	<b>Lo p-value</b>	<b>Hi Diff. Est.</b>	<b>Hi Std. Err.</b>	<b>Hi p-value</b>
<b>North Am</b>	Month 1	-9.0090	5.1865	0.0827	-14.1000	4.2469	0.0009
<b>Non-North Am</b>	Month 1	-4.0846	5.4731	0.4557	-9.9009	5.1574	0.0552
<b>North Am vs. non-North Am</b>	Month 1	4.9244	7.5323	0.5134	4.1991	6.6798	0.5297
<b>North Am</b>	Month 3	-8.7637	5.2850	0.0976	-9.8881	4.3203	0.0223
<b>Non-North Am</b>	Month 3	-2.9587	5.6708	0.6020	-15.9360	5.3021	0.0027
<b>North Am vs. non-North Am</b>	Month 3	5.8050	7.7438	0.4537	-6.0479	6.8386	0.3767

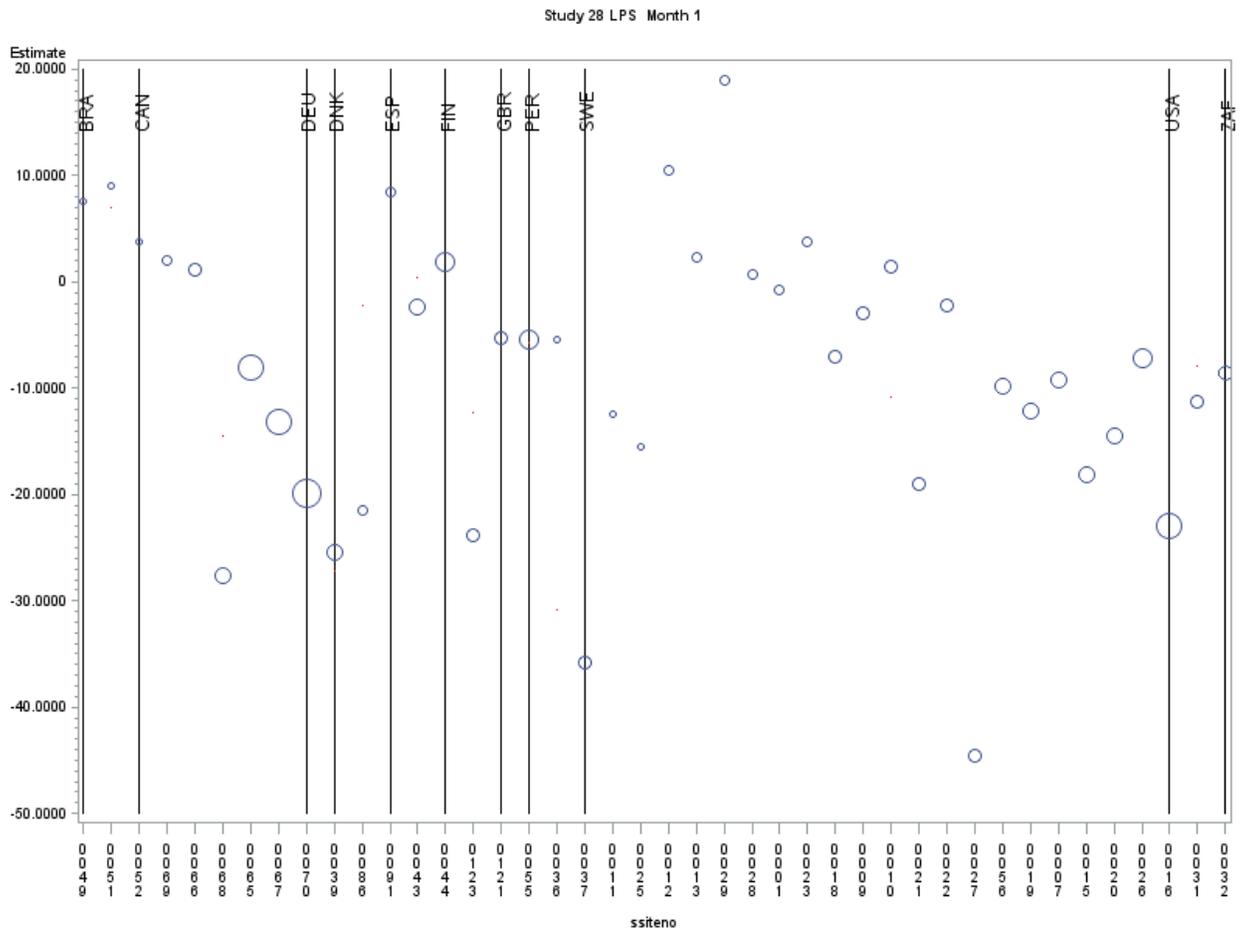
In study P06 there were two regions: North America (87%) and Japan (13%). There were significant regional differences in the 10 mg effect on WASO in the first period, with the effects in Japan being numerically in the wrong direction (Day 1: 8.6318 [26.0732 S.E.] ; Day 28: 18.3524 [25.3512 S.E.] as compared to North America: Day 1: -22.0518 [10.0736 S.E.] ; Day 28: -24.5936 [ 10.1558 S.E.]). However, the 10 mg estimated effects on LPS in the first period were consistent across the two regions(-18.9 ; -16.4 ; -19.1; -20.7 for Japan Day 1 and Day 28 and North Am. Day 1 and Day 28, respectively).

#### **4.1.5 Site Effects**

The primary analyses were adjusted for prespecified Regions rather than individual sites. However, it is important to examine the estimated treatment effects by site to see if any sites had a major impact on the results.

Figure 14 shows estimated LPS mean treatment group differences from placebo for the high dose by individual sites in study 28. The size of the plotting symbol in the figure is proportional to the number of patients randomized in the site and the sites are ordered from smallest to largest from left to right within each country. Negative differences favor the high dose group. There were 79 sites in study 28.

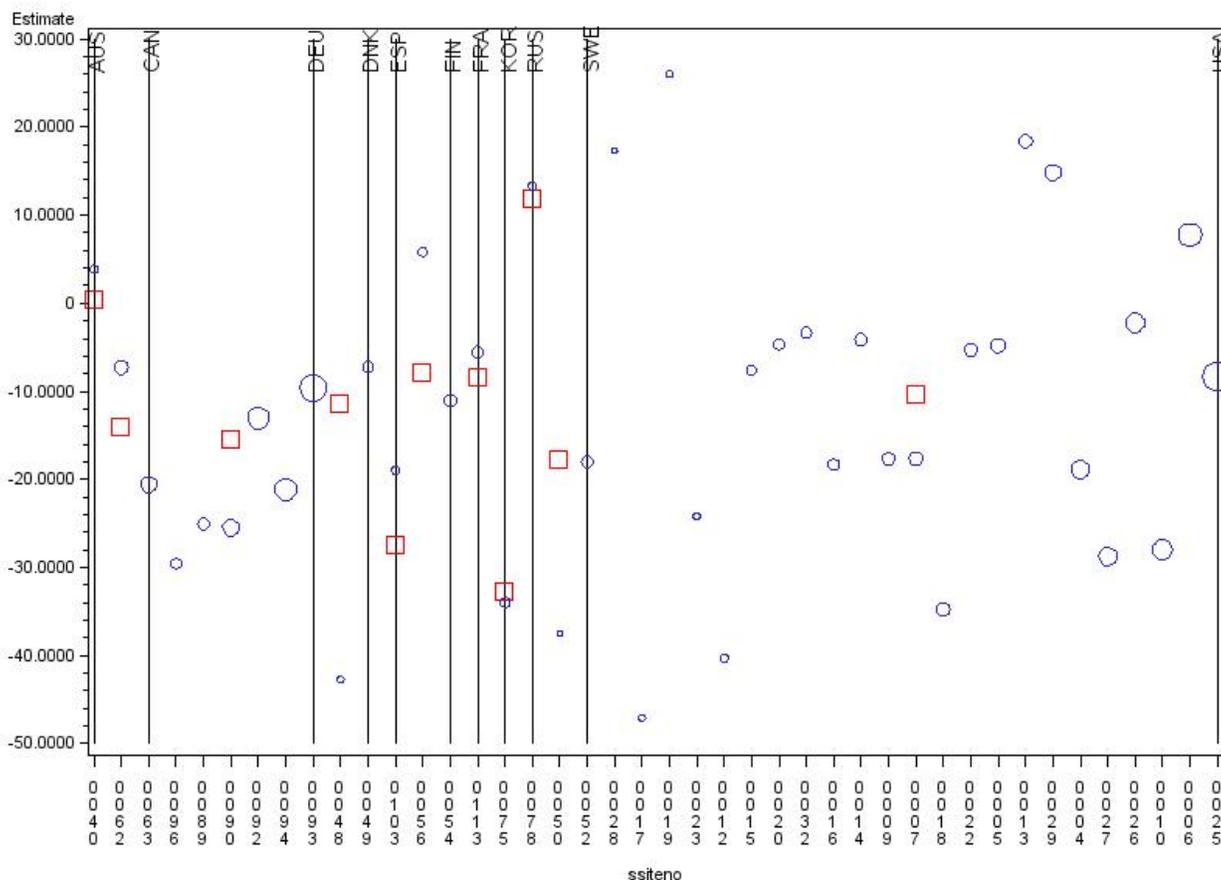
Figure 13 Study 28: Differences of Placebo from High Dose in Mean Change in LPS at Month 1 by Site



For study 28, exclusion of any one site had no effect on the significance of the high (or low) dose comparison to placebo in terms of mean change from baseline in LPS at Day 30, 90, or 1.

Figure 14 shows estimated LPS mean treatment group differences from placebo for the high dose by individual sites grouped by country in study 29. There were 90 sites overall in study 29 but only 45 were big enough to have estimates for PSG endpoints. The square plotting symbol gives the estimated difference for the country as a whole.

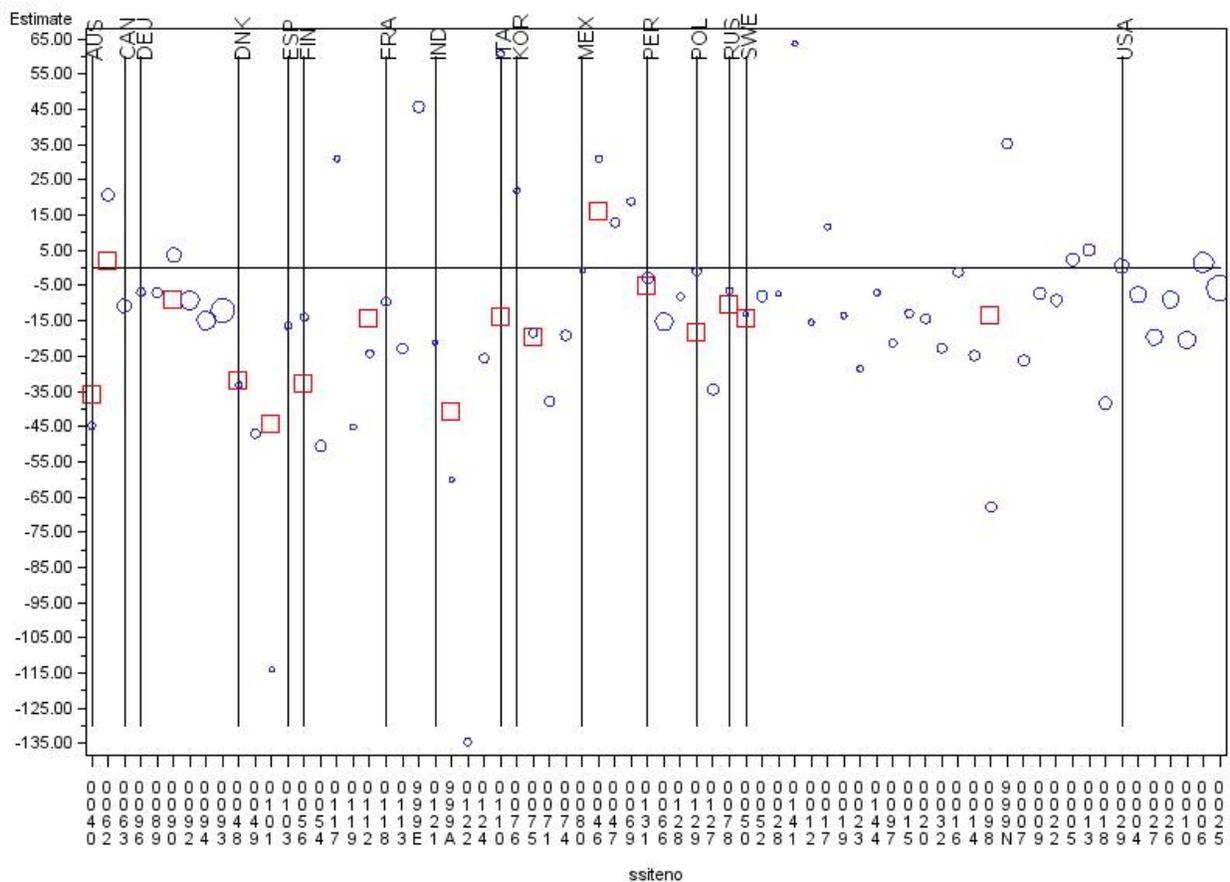
**Figure 14 Study 29: Differences of Placebo from High Dose in Mean Change in LPS at Month 1 by Site**  
P29 LPS Month 1



Exclusion of site 0004, resulted in a loss of nominal significance for the low dose ( $p=0.060$  but not for the high dose) at day 30 in terms of LPS. But this was not true for any other site for the low or high dose at day 30. Of course, as seen previously in this document, neither the low nor the high dose was significant compared to placebo in terms of mean change in LPS at day 90.

Figure 15 shows the results for the related subjective diary based endpoint, Time to Sleep Onset, at Month 1. In contrast to the objective LPS results, the high dose effect on subjective TSO was statistically significant at Month 3 (as well as at Month 1). The square plotting symbol gives the mean effect for the country.

**Figure 15 Study 29: Differences of Placebo from High Dose in Mean Change in Subj. TSO at Month 1 by Site**  
P29 sTSO 1 Month



## 4.2 Other Special/Subgroup Populations

### 4.2.1 Cohort

Cohort, Q(Questionnaire only) or PQ (Questionnaire and PSG), was a randomization stratification factor.

In study P028 24% were Q cohort (no PSG assessments). Estimated differences of high dose from placebo were numerically smaller in the Q only cohort, but the difference (interaction) was not significant for sTSTm or sTSOm.

**Table 59 Study 28: Cohort Differences for Subjective Endpoints for High Dose vs. Placebo**

<b>Endpoint</b>	<b>Contrast</b>	<b>Estimate</b>	<b>Std. Err.</b>	<b>p-value</b>
<b>sTST</b>	<b>Q vs. PQ Month 3</b>	-12.6896	9.2632	0.1711
	<b>Q vs. PQ Month 1</b>	-11.8005	8.9614	0.1882
<b>sTSO</b>	<b>Q vs. PQ Month 3</b>	6.9505	5.2149	0.1830
	<b>Q vs. PQ Month 1</b>	0.6071	5.8353	0.9172

In study P029 26% were Q cohort (no PSG assessments). The cohort differences between the high dose effects within each cohort were not significant.

**Table 60 Study 29: Cohort Differences for Subjective Endpoints for High Dose vs. Placebo**

<b>Endpoint</b>	<b>Contrast</b>	<b>Estimate</b>	<b>Std. Err.</b>	<b>p-value</b>
<b>sTST</b>	<b>Q vs. PQ Month 3</b>	-0.1235	11.1874	0.9912
	<b>Q vs. PQ Month 1</b>	-0.08504	9.7935	0.9931
<b>sTSO</b>	<b>Q vs. PQ Month 3</b>	-1.2402	7.5644	0.8698
	<b>Q vs. PQ Month 1</b>	8.4376	7.2890	0.2473

Thus, overall there was no compelling indication that the high dose effects differed substantially by cohort.

## 5 SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

As the sponsor stated in their integrated summary of efficacy patients in P006 who were randomized to 20 and 40 mg suvorexant, tended to have somewhat less severe insomnia at baseline as judged by subjective reports with longer sTSTm (>340 minutes) and shorter sTSOm values (<69 minutes) at baseline compared with the Phase 3 trials (<330 minutes and >69 minutes for sTSTm and sTSOm). With regards to PSG parameters, patients in P006 had a shorter mean WASO duration compared to the Phase 3 efficacy trials (P028 and P029). Overall, patients in the confirmatory trials (P028 and P029) had substantial impairment with regards to onset and maintenance with less than 5.5 hours of sTSTm, more than 1 hour of time to sleep onset as measured by both sTSOm and LPS and nearly 2 hours of WASO. Slightly higher severity of symptoms was observed in P029 compared to P028 as judged by the subjective assessments (sTSTm, sTSOm, and sWASOm), but not based on PSG measured sleep onset and maintenance

(WASO and LPS). Table 61 shows summary statistics for efficacy measures at Baseline by trial in the Phase 2b/3 Trials.

**Table 61 Summary Statistics for Efficacy Measures at Baseline by Trial in the Phase 2b/3 Trials**

Protocol #	Endpoint	MK-4305 LD		MK-4305 HD		Placebo	
		N	Mean	N	Mean	N	Mean
<b>Phase 2b – Dose-Finding</b>							
<b>006<sup>†</sup></b>	sTSTm	56	340.6	51	349.7	221	342.1
	sTSOm	56	68.8	51	56.7	221	62.9
	sWASOm	49	81.9	42	84.6	191	76.6
	WASO	61	97.7	59	107.3	249	100.7
	LPS	61	70.8	59	66.8	249	69.3
<b>Phase 3 – Confirmatory Efficacy</b>							
<b>028</b>	sTSTm	252	322.4	383	316.1	384	315.7
	sTSOm	252	63.3	383	68.0	384	66.9
	sWASOm	252	73.9	381	78.4	384	78.2
	WASO	193	119.2	291	117.7	290	114.9
	LPS	193	68.9	291	61.8	290	66.2
<b>029</b>	sTSTm	238	298.3	386	315.3	383	309.7
	sTSOm	238	86.0	386	74.4	383	81.3
	sWASOm	233	84.8	382	82.1	375	83.3
	WASO	150	119.6	299	119.4	295	118.4
	LPS	150	65.3	299	67.3	295	68.0
<b>Phase 3 – Long-Term Safety</b>							
<b>009</b>	sTSTm	-	-	492	319.5	245	330.0
	sTSOm	-	-	492	65.9	245	65.3
	sWASOm	-	-	488	79.6	241	71.2
	WASO	-	-	-	-	-	-
	LPS	-	-	-	-	-	-
<sup>†</sup> For Protocol 006, only data for the MK-4305 20 and 40 mg dose groups are included here since these correspond to LD and HD, respectively, for this non-elderly patient trial. MK-4305 LD = MK-4305 20 mg for patients <65 years and MK-4305 15 mg for patients ≥65 years. MK-4305 HD = MK-4305 40 mg for patients <65 years and MK-4305 30 mg for patients ≥65 years.							

Note: This table was copied from page 126 of the sponsor’s integrated summary of efficacy

Because of the potential for next day driving impairment with 40 mg as seen in the non-elderly driving study there may be interest in the lower doses. Table 62 summarizes results for the lowest dose in each study (10 mg in the non-elderly adult study 06 and 15 mg for the elderly in studies 28 and 29). The p-values shown should be regarded with caution since they are not adjusted for multiplicity and the low dose was always relegated by the sponsor to secondary status.

**Table 62 Summary of Efficacy Measures for Lowest Doses in Phase 2B/3 studies**

Study	Dose	Measure	Time	Baseline Placebo	Placebo Mean change estimate	Baseline MK4305	Diff Estimate	Diff StdErr	Pvalue
6	10 <sup>#</sup>	LPS	Night 1	67.3	-12	65.8	-19.1	8.14	0.0201
28/29	15*	LPS	Night 1	67.5	-14.9	66.5	-10.2	3.75	0.0068
6	10 <sup>#</sup>	sTSO	Week 1	61.7	-1.89	65.5	-4.03	4.7	0.3919
28/29	15*	sTSO	Week 1	69.4	-7.59	65.9	-6.66	2.81	0.0176
6	10 <sup>#</sup>	WASO	Night 1	98.7	-21	108	-18.1	9.36	0.0549
28/29	15*	WASO	Night 1	127	-2.66	133	-39.3	4.07	<.0001
6	10 <sup>#</sup>	LPS	Night 28	67.3	-16.7	65.8	-20.2	8.5	0.0185
28/29	15*	LPS	Night 28	67.5	-25.3	66.5	-5.2	3.5	0.1335
6	10 <sup>#</sup>	sTSO	Week 4	61.7	-3.3	65.5	-10.5	4.7	0.0254
28/29	15*	sTSO	Week 4	69.4	-12.5	65.9	-3.8	3.5	0.2776
6	10 <sup>#</sup>	WASO	Night 28	98.7	-23.4	107.8	-18.8	9.5	0.0481
28/29	15*	WASO	Night 28	127.1	-1.0	133.5	-27.0	4.6	<.0001

<sup>#</sup> Non-elderly only in study 6

\* Elderly only dose

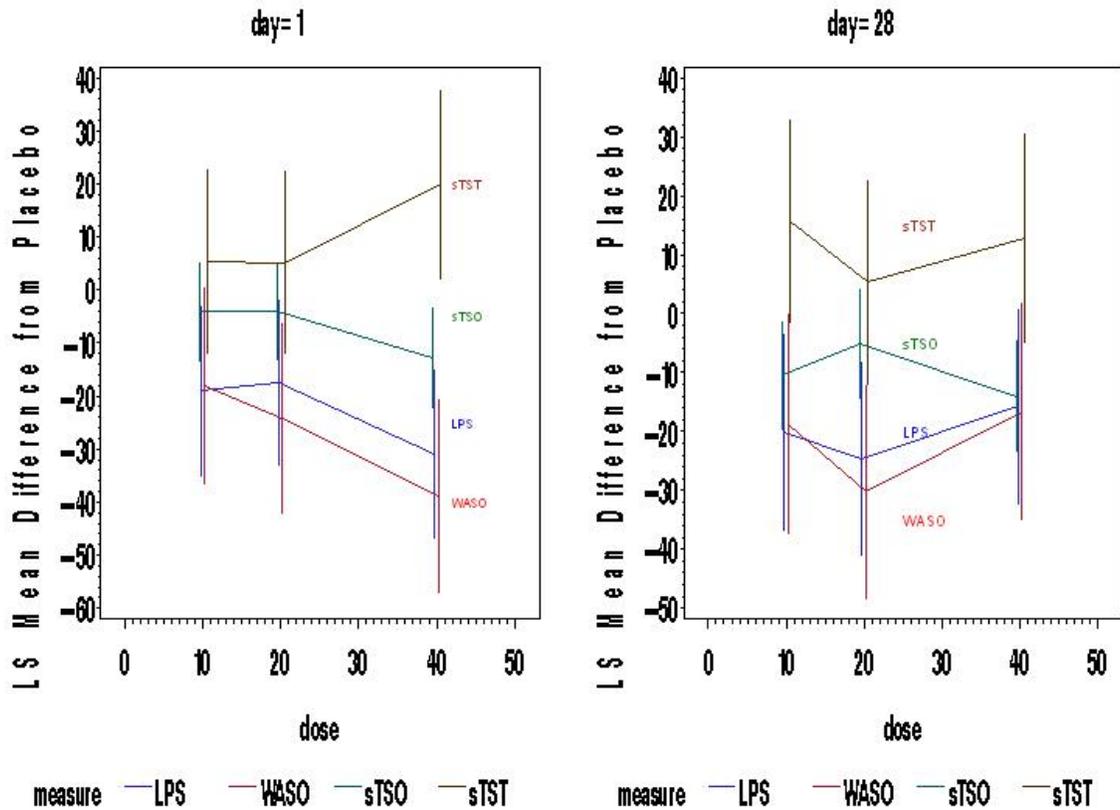
An effect on Latency to Persistent Sleep as measured by Polysomnography was replicated for the Month 1 visit, the primary timepoint, but not at the final timepoint of Month 3. However, the

corresponding subjectively measured item, Time to Sleep Onset, as determined from weekly averages of subjects' sleep ratings captured in diaries was statistically significant at Month 1 and Month 3 in both studies. The failure of LPS at Month 3 in study 29 means any secondary efficacy measures of Sleep Onset must be viewed as exploratory at best.

Some sponsors of investigational sleep drugs design separate studies for elderly and non-elderly and/or for demonstration of sleep maintenance and sleep onset effects. The Suvorexant phase 3 studies were ambitious in that they aimed to demonstrate effects for both elderly and non-elderly patients on both Sleep Maintenance and Onset, in terms of both an objective and a subjective assessment for each in the same study. They also put more weight on the high dose (the low dose had 30 to 40% less patients by design) and the multiplicity adjustment method tested the high dose first, such that the low dose could only be tested if the high dose was first significant at multiple timepoints on both subjective and objective assessments. Also, a Bonferroni adjustment was made so that Maintenance endpoints could be tested regardless of the significance of Onset endpoints or vice versa, but at the cost of the tests being done at the significance level of 0.025 rather than 0.05. There was also a Hochberg adjustment made because of testing both objective and subjective endpoints such that if, for example, the subjective onset endpoint was not significant at 0.025 then the objective onset endpoint needed to be significant at 0.0125. Under these circumstances the low dose was not statistically significant at the multiplicity adjusted level in study 29 at Month 1 or Month 3 for either of the primary sleep onset endpoints (objective LPS or subjective time to sleep onset) or for sTSO in study 28. Therefore, the statistically significant Onset effects of the low dose (15 mg for elderly and 20 mg for non-elderly) observed in study 28 for LPS were not strictly replicated in study 29. The low dose p-values in study 28 were 0.0004 and 0.0061 for LPS at month 1 and month 3 and 0.0519 and 0.0377 at month 1 and month 3 for sTSO. In study 29 the low dose p-values were 0.0306 and 0.9322 for LPS at month 1 and month 3 and 0.0498 and .0389 for sTSO at month 1 and month 3. The low dose was statistically significant at Month 1 and Month 3 for the primary maintenance endpoints in both study 28 and study 29.

**Figure 16** summarizes dose response at day 1 on the left and day 28 on the right for study 6 in terms of the key efficacy measures used later in the phase 3 studies.

Figure 16 Study 06: Dose Response in First Period for Various Efficacy Measures



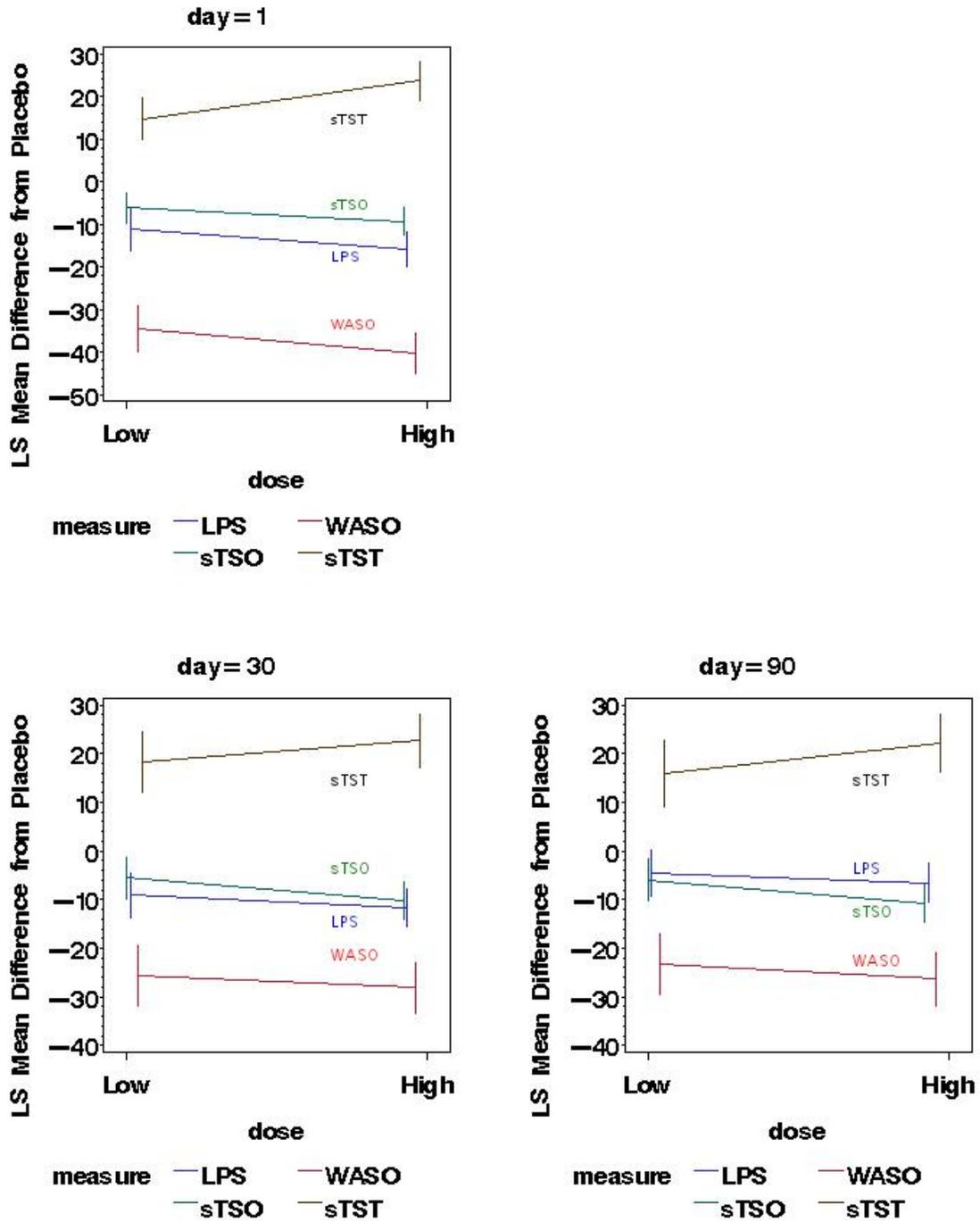
An exploratory test for linear trend among the doses 10 through 40 in the first period of study 06 was not significant for sTST (.49,  $p=0.175$ ), sTSO (-.35,  $p=0.063$ ), or LPS (-.43,  $p=0.064$ ), but was for WASO (-.71,  $p=0.019$ ) at day 1.

An exploratory assessment for a linear trend among the doses 10 through 40 in the first period was not significant for sTST (.03,  $p=0.941$ ), sTSO (-.21,  $p=.277$ ), LPS (0.22,  $p=0.308$ ), or WASO (0.31,  $p=0.362$ ) at day 28. The trend in WASO was not significant at day 28 even if the 80 dose was included (-.156,  $p=.225$ ).

Including 80 mg data in fact did not change the nominal significance of any of the estimated linear trends. It should be noted that this was a small study though, so the power was not too high for detecting such a trend and only the first period data was used for this exploratory analysis because of the treatment by period interaction found for LPS. Overall, it seems that the dose response is uncertain based on study 06.

Figure 17 shows exploratory pooled phase 3 data results for the primary efficacy measures at each time point.

Figure 17 Study 28/29 Pooled Estimates for Primary Efficacy Measures



## 5.2 Conclusions and Recommendations

The clinical trial efficacy data provided in this application seems to clearly support the efficacy of Suvorexant for Sleep Maintenance. In the application there are two similarly designed 3 month placebo controlled phase 3 studies and one early phase 2B dose finding crossover study. The evidence for an effect on Sleep Onset was weaker than that for maintenance. In one study at the 3 month visit night the effect on latency as measured objectively by Polysomnography did not achieve statistical significance, thus failing to replicate the statistically significant effect demonstrated in the other study. However, the high dose effect at Month 1 was significant in both studies and there was replication of the effect on the corresponding subjective assessment, the patient reported weekly average of the Time to Sleep Onset, at Month 3 (as well as Month 1).

The Suvorexant phase 3 studies were ambitious in that they aimed to demonstrate effects for both elderly and non-elderly patients on both Sleep Maintenance and Onset, in terms of both an objective and a subjective assessment for each, in the same study. They also put more weight on the high dose (the low dose had 30 to 40% less patients by design) and the multiplicity adjustment method tested the high dose first, such that the low dose could only be tested if the high dose was first significant at multiple timepoints on both subjective and objective assessments. After taking into account the prespecified adjustments for multiple testing the low dose was only statistically significant for objective latency to persistent sleep in study 28 (not significant for objective latency in study 29 or for subjective time to sleep onset in study 28 or study 29). However, the low dose was statistically significant for objective and subjective primary endpoints for maintenance in both of these studies.

In the non-elderly next day driving safety study the 40 mg dose had significant asymmetry of differences from placebo with significantly more being positive and higher than the impairment threshold of interest (2.40) on both days 2 and 9. The low dose also was significant on day 2 in the symmetry analysis but not on day 9. There was also some other evidence that the driving effect might be dose related which would suggest a lower dose if it was still efficacious.

In a prior phase 2B crossover study in non-elderly adults a lower dose, 10 mg, as well as a higher dose, 80 mg, were studied in addition to the adult doses used later in Phase 3. There was a suggestion of efficacy of 10 mg, particularly for wake time after sleep onset, based on this study data but there is no existing means of replication for the 10 mg dose and no 10 mg data for the elderly. If the phase 3 doses are considered to have too much of a risk of next day driving impairment then in this reviewer's opinion another study may be needed.

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

TRISTAN S MASSIE  
05/07/2013

KUN JIN  
05/07/2013  
I concur with the review.

KOOROS MAHJOOB  
05/07/2013  
I concur with the review.



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Science  
Office of Biostatistics

# Statistical Review and Evaluation

## CLINICAL STUDIES

**NDA/Serial Number:** 204-569  
**Drug Name:** Suvorexant (MK-4305)  
**Indication:** Insomnia  
**Study number:** Study P025  
**Applicant:** MERCK SHARP & DOHME CORP., A SUBSIDIARY OF  
MERCK & CO., INC.  
**Date(s):** Filing Letter date: 11/09/2012  
PDUFA date: 06/29/2013  
Completion date: 05/01/2013  
**Review Priority:** S  
**Biometrics Division:** DB VI  
**Statistical Reviewer:** Ling Chen, Ph.D., Mathematical Statistician, Special Projects  
Team, DBVI/OB/OTS  
**Concurring Reviewers:** Stella Machado, Ph.D., Division Director, and Acting Team  
Leader, DBVI/OB/OTS  
**Medical Division:** Controlled Substance Staff  
**The CSS Team:** Chad J. Reissig, Ph.D., Pharmacologist, OD/CSS  
Silvia Calderon, Term leader, OD/CSS  
**Project Manager:** Sandra Saltz, OD/CSS  
**Keywords:** Crossover design; Drug abuse potential study; Self-reported endpoint;  
Multiple endpoints

## Table of Contents

<b>LIST OF TABLES.....</b>	<b>3</b>
<b>LIST OF FIGURES.....</b>	<b>3</b>
<b>2. REVIEW REPORT ON STUDY P025.....</b>	<b>5</b>
2.1 OVERVIEW.....	5
2.1.1 Objectives of the study.....	5
2.1.2 Study design.....	5
2.1.3 Abuse potential measures and data collection times.....	6
2.1.4 Subject and Subject Disposition.....	7
2.1.5 Statistical methodologies used in the Sponsor's analyses.....	7
2.1.5 Sponsor's Summary and Conclusions.....	8
2.2 DATA LOCATION.....	9
2.3 REVIEWER'S ASSESSMENT.....	9
2.3.1 Missing data issue.....	9
2.3.2 Primary Analysis.....	11
2.3.2.1 Descriptive Statistics.....	11
2.3.2.2 Statistical Testing.....	14
2.3.3 Secondary Analysis.....	15
<b>3. CONCLUSION.....</b>	<b>20</b>
<b>4. REFERENCES.....</b>	<b>21</b>
5.1 APPENDIX I.....	22
IN THE CASE THAT ONE OF THE ITEMS WAS MISSING THE RELATED SCORE WAS NOT CALCULATED.....	22
5.2 APPENDIX II.....	23
5.3 APPENDIX III.....	27

**List of Tables**

Table 1: Summary statistics for Emax of Drug Liking VAS ..... 11  
 Table 2: P-values of the significance tests for Emax of Drug Liking VAS (n=32)..... 15  
 Table 3: Summary of the test results for Emax of Bowdle VAS..... 15  
 Table 4: Summary statistics and testing results for Emax of Bad Effects VAS..... 16  
 Table 5: Summary statistics and testing results for Emax of Bowdle External Perception..... 17  
 Table 6: Summary statistics and testing results for Emax of Bowdle Internal Perception..... 17  
 Table 7: Summary statistics and testing results for Emax of Good Effects VAS ..... 18  
 Table 8: Summary statistics and testing results for Emax of High VAS..... 18  
 Table 9: Summary statistics and testing results for Emax of Overall Drug Liking VAS..... 19  
 Table 10: Summary statistics and testing results for Emax of Take Drug Again VAS..... 19  
 Table 11: Summary of the results from significance tests for the abuse potential measured  
 considered in this review ..... 20  
 Table 12: P-values of the significance tests for Emax of secondary measures excluding Bowdle  
 External Perception and Bowdle Internal Perception (n=32) ..... 27

**List of Figures**

Figure 1: Individual time course response profiles for Drug Liking VAS (Z30) ..... 10  
 Figure 2: Boxplots for five treatments and the differences between suvorexant and zolpidem for  
 Drug Liking VAS (N=32) ..... 12  
 Figure 3: Mean and median response curves for suvorexant and zolpidem for Drug Liking VAS12  
 Figure 4: Dose response curves of suvorexant in difference from zolpidem ..... 13  
 Figure 5: Mean time course profiles for Drug Liking VAS ..... 13  
 Figure 6: Heat map for Emax of Drug Liking VAS by treatment by subject ..... 14  
 Figure 7: Mean time course profiles for Bad Effects VAS ..... 23  
 Figure 8: Mean time course profiles for Bowdle External Perception ..... 23  
 Figure 9: Mean time course profiles for Bowdle Internal Perception ..... 24  
 Figure 10: Mean time course profiles for Good Effects VAS ..... 24  
 Figure 11: Mean time course profiles for High VAS ..... 25  
 Figure 12: Overall Drug Liking VAS at hours 12 and 24 ..... 25  
 Figure 13: Take Drug Again VAS at hours 12 and 24 ..... 26

## 1. Executive Summary

Study P025 was a randomized, double-blind, placebo- and active comparator controlled 6-way crossover study to evaluate the abuse potential of single doses (40 mg, 80 mg and 150 mg) of suvorexant compared to placebo and 2 doses (15 mg and 30 mg) of zolpidem in healthy male and female recreational polydrug users.

After the Qualification Session, 36 subjects were randomized to the Treatment Session, and 32 subjects completed the study.

The primary objectives were to assess the relative abuse liability of suvorexant compared to placebo in recreational polydrug users, as measured by Drug Liking bipolar visual analogue scale (VAS), and to confirm the abuse liability of zolpidem compared to placebo as measured by the Drug Liking VAS, in order to confirm study validity. The validation test was successful for both doses of zolpidem. However, the study showed that the relative abuse liability of suvorexant was greater than that of placebo.

The comparison between suvorexant and zolpidem for the primary measure Drug liking VAS was considered as one of the secondary objectives in the Sponsor's analysis. The study did not demonstrate the lower abuse potential of suvorexant compared to zolpidem for the primary measure.

Per CSS request, the reviewer also studied 7 secondary abuse potential measures: Bad Effects VAS, Bowdle External Perception, Bowdle Internal Perception, Good Effects VAS, High VAS, Overall Drug Liking VAS, and Take Drug Again VAS. Suvorexant shows greater abuse potential than placebo for any of these measures. Except for Bad Effects VAS, in 10 out of 36 comparisons, the responses to suvorexant were significantly less than those of zolpidem (See [Table 11](#)). The study results also showed that suvorexant 40 mg and 150 mg had significantly lower bad effects than zolpidem 30 mg.

Some data were missing at early hours due to adverse experiences associated with treatment that prevented these subjects from performing the scheduled tests. Approximately 21.9% (7/32) of subjects had missing data at hour 0.5, of which six were males. The missing data were not imputed by either the Sponsor or the reviewer. The reason for not imputing missing data can be found in [2.3.1 Missing data issue](#).

In summary, suvorexant has higher abuse liability than placebo, and does not demonstrate lower abuse liability than zolpidem for the primary measure, Drug Liking VAS. Considering the results from both the reviewer's primary and secondary analyses, and the fact that there is no apparent dose-response for suvorexant, the reviewer concludes that the abuse potential of suvorexant may be similar to or lower than zolpidem.

## 2. Review Report on Study P025

### 2.1 Overview

#### 2.1.1 Objectives of the study

##### Primary objectives

- To assess the relative abuse liability of suvorexant compared to placebo in recreational polydrug users, as measured by Drug Liking visual analogue scale (VAS).
- To confirm the abuse liability of zolpidem compared to placebo as measured by the Drug Liking VAS, in order to confirm study validity.

##### Secondary objectives

- To assess the relative abuse liability of suvorexant compared to zolpidem, as assessed by the Drug Liking VAS.
- To assess the relative abuse liability of suvorexant compared to zolpidem and placebo as assessed by measures of positive effects (e.g., Drug Liking, Take Drug Again, High, Good Effects, Overall Drug Liking VASs and Morphine Benzodrine Group (MBG) subscale and Benzodrine Group (BG) of the Addiction Research Center Inventory (ARCI).
- To assess the relative abuse liability of suvorexant compared to zolpidem and placebo as assessed by measures of negative effects (e.g., Lysergic Acid Diethylamide (LSD) subscale of the ARCI, and Bad Effects VAS).
- To evaluate the relative abuse liability dose response profile of suvorexant (40, 80, and 150 mg) compared to zolpidem (15 mg and 30 mg) relative to placebo as measured by the primary and secondary variables.
- To further characterize the abuse liability of suvorexant compared to placebo and zolpidem as measured by items from the Subjective Effects VASs and ARCI scales not listed above, and the Drug Similarity VASs.
- To assess the cognitive effects of suvorexant compared to placebo in recreational polydrug users, as measured by the Choice Reaction Time (CRT) test, the Divided Attention test (DAT), the Sternberg Short-Term Memory Test (STM) and the Hopkins Verbal Learning Test Revised (HVLTR).
- To evaluate the safety and tolerability of single oral doses of suvorexant up to 150 mg.

*Reviewer's comment: Even though the comparison between suvorexant and zolpidem for the primary measure Drug Liking VAS was considered as one of the secondary objectives, this comparison is considered as primary in this reviewer's analysis.*

#### 2.1.2 Study design

The design was a randomized, double-blind, placebo- and active comparator controlled 6-way crossover study to evaluate the abuse potential of single doses of suvorexant compared to placebo and 2 doses of zolpidem in healthy male and female recreational polydrug users. Subjects participated in a Screening Visit, Qualification Session, 6 Treatment Sessions, and a safety Post-Study Visit.

Within 28 days of a standard medical screening, subjects attended a randomized double-blind, Qualification Session (Part I) in which they received either zolpidem 20 mg or placebo in a crossover manner, each separated by approximately 24 hours, to ensure that they could discriminate and show positive effects of the active comparator. Each treatment was followed by serial PD measurements to assess abuse potential. Subjects who had a clinically significant (as judged by the investigator or designee) positive urine drug screen at admission to the Qualification Session were not dosed but could be rescheduled for another session at the discretion of the investigator (provided subsequent drug screen was negative). Following qualification, eligible subjects who passed the Qualification Session were enrolled and randomized to the Treatment Session (Part II). Drug administration occurred on Day 1 of each Treatment Visit followed by PD, pharmacokinetic (PK) and safety assessments conducted for up to 24 hours post dose.

Subjects received each of the following 6 treatments in a randomized, double-blinded, balanced fashion (one per Treatment Period):

1. Placebo: 5 suvorexant placebo (3 x 30 mg + 2 x 10 mg suvorexant placebo) and 6 zolpidem placebo
2. zolpidem 15 mg: 3 x 5 mg zolpidem tablets and 10 placebo tablets: (3 x 5 mg zolpidem placebo and 5 x 30 mg + 2 x 10 mg suvorexant placebo)
3. zolpidem 30 mg: 6 x 5 mg zolpidem tablets and 7 placebo suvorexant tablets (5 x 30 mg + 2 x 10 mg suvorexant placebo tablets)
4. suvorexant 40 mg: 1 x 30 mg and 1x10 mg suvorexant and 11 placebo tablets (4 x 30 mg + 1 x 10 mg suvorexant placebo and 6 zolpidem placebo)
5. suvorexant 80 mg: 2 x 30 mg and 2 x 10 mg suvorexant and 9 placebo tablets (3x 30 mg suvorexant placebo and 6 zolpidem placebo)
6. suvorexant 150 mg: 5 x 30 mg suvorexant tablets and 8 placebo tablets: (2 x 10 mg suvorexant placebo and 6 zolpidem placebo)

Treatment Periods were separated by at least 10 days to minimize potential carry-over effects. Subjects were administered treatment in each period provided that they showed no clinical evidence of persistent drug effect of any study drugs. Subjects who had a clinically significant (positive) urine drug screen (as judged by the investigator or designee) prior to any of the Treatment Periods were not dosed but could be rescheduled for another session at the discretion of the investigator (provided subsequent drug screen was negative).

### **2.1.3 Abuse potential measures and data collection times**

Besides the primary measure Drug Liking VAS, the following measures were also evaluated in the study:

Positive Effects: High VAS, Overall Drug liking VAS, Take Drug Again VAS, Good Effects VAS, ARCI MBG and Subjective Drug Value VAS

Negative Effects: ARCI LSD and Bad Effects VAS

Other Effects Measures: Alertness/Drowsiness VAS, Any Effect VAS, ARCI PCAG, ARCI BG, Overall Familiarity, and Bowdle VAS

Data were collected at hours 0.5, 1, 2, 3, 4, 6, 8, 12, and 24 postdose for Drug Liking VAS, Good Effects VAS and Any Effects VAS. For High VAS, Alertness/Drowsiness, and Bowdle VAS and

ARCI Scales, pre-dose data were also collected. For Subjective Drug Value, Overall Drug Liking VAS and Take Drug Again VAS, data were collected at hours 12 and 24. For Drug Similarity VAS, data were collected only at hour 12.

The study also collected data for Choice Reaction Time (CRT), Divided Attention Test (DAT), and Sternberg Short Term Memory (STM) Test at the same time points as High VAS, and Hopkins Verbal Learning Test Revised –Recall (HVLT-R) at hours 0.5, 1.5, 3, 6, and 24 postdose.

### 2.1.4 Subject and Subject Disposition

	<u>Part I Qualification</u> <u>Session</u>	<u>Part II Treatment</u> <u>Session</u>
RANDOMIZED:	73	36
Male (age range)	38 (20-53 yr)	19 (23-53 yr)
Female (age range)	35 (21-53 yr)	17 (22-48 yr)
COMPLETED <sup>†</sup> :	36	32
DISCONTINUED <sup>†</sup> :	37	4
Clinical adverse experience	3 <sup>‡</sup>	0
Laboratory adverse experience	0	0
Withdrew consent	5	3
Other	29 <sup>§</sup>	1 <sup>¶</sup>

<sup>†</sup> ‘Completed’ represents the total number of subjects who completed both periods of Part I regardless of whether they qualified for Part II. ‘Discontinued’ represents the total number of subjects that either did not complete Part I or did not continue on to Part II.

<sup>‡</sup> AN 0011 discontinued due to AE of vomiting, AN 0022 discontinued due to AE of elevated blood pressure, and AN 0024 discontinued due to AE of emesis.

<sup>§</sup> Nineteen (19) subjects failed qualification and another 10 qualified subjects were discontinued after Part I as enrollment requirements in Part II were already met by other subjects.

<sup>¶</sup> One (1) subject was discontinued due to scheduling conflicts.

Note: This table is from page 4 of the Sponsor’s report.

### 2.1.5 Statistical methodologies used in the Sponsor’s analyses

The primary endpoint was the peak effect (maximum score recorded between 0.5 and 12 hours post dose) for the Drug Liking VAS. To address the study goals, individual peak effect values were evaluated with a linear mixed effects model containing fixed effects for treatment, gender, first-order carryover, and period, and subject as a random effect. Carryover was tested at the 0.25 level and dropped from the model if found to be not significant. If the upper bound of the 95% confidence interval of the mean peak effect treatment difference (suvorexant - placebo) was less than 15 on the VAS-drug liking scale for each suvorexant dose, then the first primary hypothesis (single doses of suvorexant had abuse potential no greater than placebo) would be supported. The second primary hypothesis that single doses of zolpidem had greater abuse liability than placebo would be supported if the lower bound of the 95% confidence interval of the mean peak effect treatment difference (zolpidem - placebo) was greater than zero for each zolpidem dose. The secondary hypothesis: that single doses of suvorexant had abuse liability less than each dose of

zolpidem will have been supported if the upper bound of the 95% confidence interval of the mean peak effect treatment difference (suvorexant - zolpidem) was less than zero for each dose. Means and 95% confidence intervals were also calculated for each treatment. Secondary endpoints (other VAS measures, ARCI, drug similarity, SDV) were evaluated in a similar manner. Summary statistics were provided to explore the cognitive and psychomotor effects of suvorexant in healthy subjects, e.g. Choice Reaction Time (CRT) test, Divided Attention test (DAT), Sternberg Short-Term Memory test (STM), and Hopkins Verbal Learning Test-Revised (HVLTR).

*Reviewer's comments: In this study, the Sponsor wanted to show that suvorexant has no abuse potential. This can be seen from the primary objective of the study. The statistical method used in the study was that if the upper bound of the 95% confidence interval of the mean peak effect treatment difference (suvorexant - placebo) was less than 15 on the VAS-drug liking scale for each suvorexant dose, then the first primary hypothesis (single doses of suvorexant had abuse potential no greater than placebo) would be supported. The recently proposed criterion for this comparison by Chen and Bonson (2013) is to use an equivalence test with a margin 11. This is equivalent to comparing the upper bound of the 95% confidence interval of the difference between suvorexant and placebo to 11 instead of 15. The new method proposed by Chen and Bonson is more stringent than the sponsor's criterion. In other words, if the comparison fails in Chen and Bonson's criterion, it may not fail by using this Sponsor's criterion. Nevertheless, the primary comparison failed even by using the Sponsor's own criterion (See next section).*

*Please notice that the Sponsor did not check model assumptions in their analyses.*

## **2.1.5 Sponsor's Summary and Conclusions**

### Summary of the study results

- Suvorexant shows greater abuse potential than placebo in recreational polydrug users, as measured by the Drug Liking VAS.
- Zolpidem demonstrates greater abuse potential than placebo, as measured by the Drug Liking VAS.
- Suvorexant shows similar abuse potential as zolpidem as measured by the Drug Liking VAS
- Both suvorexant and zolpidem showed greater abuse potential than placebo on other positive measures of drug abuse potential.
- The effects of suvorexant and zolpidem were generally similar on other positive measures of drug abuse potential. However, on High VAS, ARCI MBG subscale and Bowdle VAS, all doses of suvorexant showed statistically significantly less effect than zolpidem 30 mg.
- There was no apparent dose-response for suvorexant on positive measures of drug abuse potential, whereas higher dose (30 mg) of zolpidem appeared to have greater effects than the low dose (15 mg) on most measures.
- Both suvorexant and zolpidem showed statistically more negative effects than placebo as measured by "bad effect VAS" and ARCI LSD subscale (measuring dysphoria effects).
- Suvorexant and zolpidem demonstrated comparable pharmacological effects and impairment on psychomotor performance at doses evaluated in this study.
- Suvorexant (40, 80 and 150 mg) and zolpidem (15 and 30 mg) are generally well tolerated in recreational polydrug users following single dose administration. The incidence of abuse potential AEs was generally lower following administration of suvorexant than zolpidem.

## Conclusions

Suvorexant shows greater abuse potential than placebo as measured by subjective abuse potential measures in recreational polydrug users. Although suvorexant showed similar effect as zolpidem on "Drug Liking VAS", the fact that there was a relatively flat dose-response for suvorexant on most measures of abuse potential (whereas zolpidem tended to have greater effects at higher doses), less effect with suvorexant than high dose zolpidem on some of the secondary measurements (e.g. High VAS, ARCIMBG and Bowdle VAS), and lower incidence of abuse potential AEs with suvorexant suggested that suvorexant may have overall less abuse potential than zolpidem.

### **2.2 Data Location**

The analysis dataset is located at

<\\cdsesub1\EVSPROD\NDA204569\0011\m5\datasets\p025\analysis\datasets\analds25.xpt>

### **2.3 Reviewer's Assessment**

In the reviewer's report P, S40, S80, S150, Z15, Z30 denote placebo, suvorexant 40 mg, 80 mg and 150 mg, and zolpidem 15 mg and 30 mg, respectively.

The statistical analysis was based on the data from 32 completers. We define the primary endpoint Emax to be the maximum response or maximum change from predose response during 8 hours after dosing. This is different from the Emax calculated by the Sponsor. This Sponsor defined Emax to be the maximum response of maximum change from predose response during 12 hours after dosing.

Per CSS request besides the primary measure Drug Liking VAS, the reviewer's secondary analysis included High VAS, Good Effects VAS, Bad Effects VAS, Overall Drug Liking VAS, Take Drug Again VAS as well as two measures related to Bowdle VAS (Bowdle External Perception, and Bowdle Internal Perception), which were measured on the log<sub>10</sub> scale.

Because all doses of suvorexant and zolpidem showed significantly higher responses than placebo in both the reviewer's analyses and the Sponsor's analyses for these measures considered in this review report, the details for the comparisons with the placebo are not presented in this report.

#### **2.3.1 Missing data issue**

The reviewer examined the data for abuse potential measures using heat map displays proposed by Chen and Wang (2012).

Figure 1 shows the individual time course response profile for zolpidem 30 mg for Drug Liking VAS. The orange line separates the responses by gender. The subjects above the orange line are females, and the subjects below the orange line are males. Colors blue, white, and red denote dislike, neutral and like, respectively. The grey color indicates missing data. At hour 0.5, 21.9% (7/32) of subjects have missing data. Only one is female. Subject #5 has missing data at hours 1,

2, and 3. The reviewer requested the Sponsor to explain the reasons for the missing data. The Sponsor responds that:

*For the primary endpoint of the study, “drug liking” VAS, there are 12 subjects that have some missing data. ... All 12 subjects are missing data because of adverse experiences associated with treatment that prohibited them from performing the scheduled tests.*

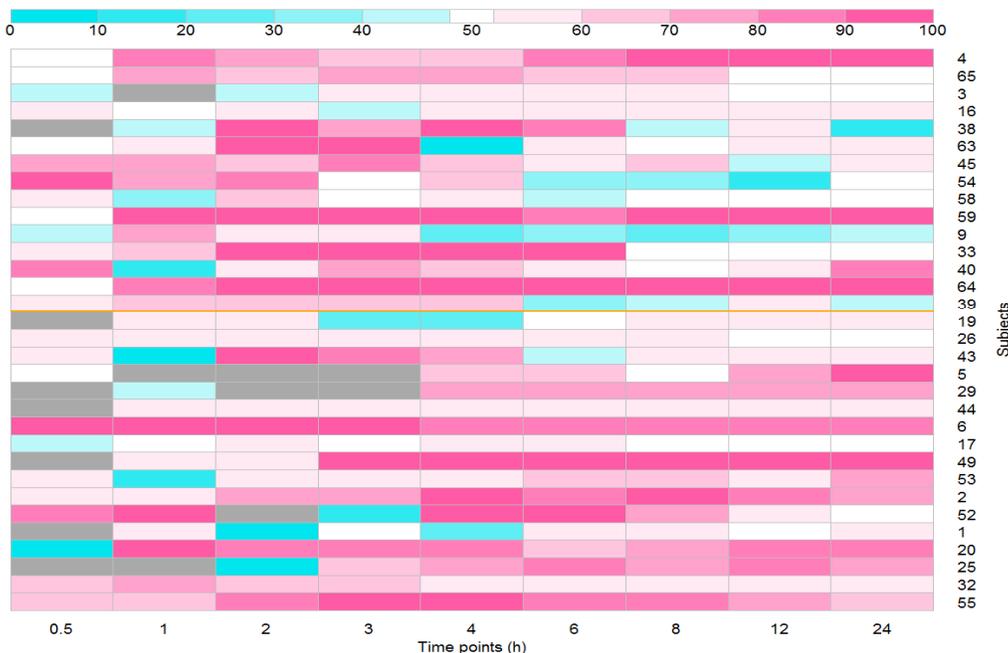
The Sponsor also states that

*... N=30 subjects are missing at least one data point from the PD endpoints. There are 2722 data points missing out of 83,230 records; about 3.3% of the data.*

The information about the Sponsor’s responses can be found at

<\\cdsesub1\EVSPROD\NDA204569\0016\m1\us\efficacy-information-amendment-18mar2013.pdf>

Notice that if missing data occurred at a time point for a study subject for Drug Liking VAS, in most cases, the missing data also occurred at the same time point for the subject for other abuse potential measures, because due to the AE, the subject was unable to answer questions.



**Figure 1: Individual time course response profiles for Drug Liking VAS (Z30)**

The missing data in this study are “informative missing,” meaning that we know the reason for the missing. These missing data cannot be imputed for the following reasons:

- When the data are collected at multiple time points. The intention is to collect the information from subjects at the moment. Thus, either the last observation carried forward or the first observation after the missing observation moved backward does not make sense for this study. In addition, the primary endpoint is Emax. Using existing data of a subject to impute the subject’s missing data will not change Emax.

- Another way may be to impute the missing data at a particular time point by the average of the responses from other subjects who have data at the time point. However, this method will violate the principle of the crossover design, that is, using self-response as the control.

Even though, statistically, missing data can be imputed in many cases, it is definitely not the case for such a study. Therefore, missing data were not imputed in the reviewer's analyses. However, the reviewer suggests the Sponsor to improve their selection of subjects in the Qualification Session in future studies.

### 2.3.2 Primary Analysis

#### 2.3.2.1 Descriptive Statistics

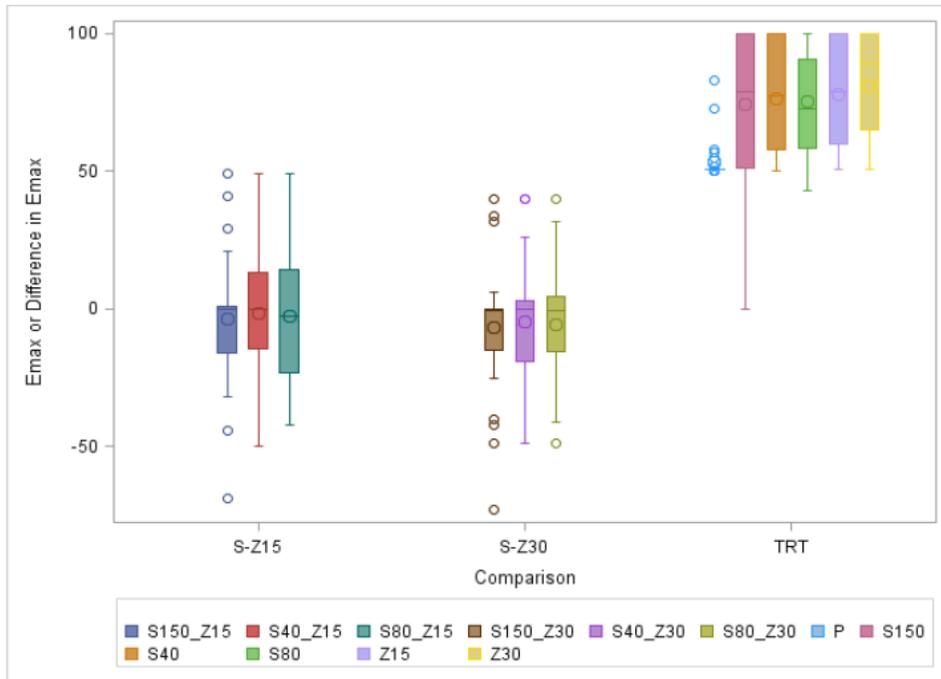
Table 1 summarizes the mean, standard error, minimum, the first quartile (Q<sub>1</sub>), median, the third quartile (Q<sub>3</sub>), and maximum for five treatments in the study and for the treatment differences between suvorexant and zolpidem for Emax of Drug Liking VAS.

**Table 1: Summary statistics for Emax of Drug Liking VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	53.09	1.22	50	51	51	51	83.00
S40	32	76.28	3.47	50	76.28	77.5	100	100.00
S80	32	75.28	3.25	43	75.28	73	91	100.00
S150	32	74.16	4.26	0	74.16	79	100	100.00
Z15	32	77.91	3.45	51	60	79	100	100.00
Z30	32	80.94	3.21	51	65	89	100	100.00
S40_Z15	32	-1.63	4.40	-50	-14.5	0	13.5	49.00
S40_Z30	32	-4.66	3.67	-49	-19	0	3	40.00
S80_Z15	32	-1.63	4.20	-42	-23	-2.5	14.5	49.00
S80_Z30	32	-5.66	3.48	-49	-15.5	-0.5	4.5	40.00
S150_Z15	32	-3.75	4.02	-69	-16	0	1	49.00
S150_Z30	32	-6.78	3.98	-73	-15	-0.5	0	40.00

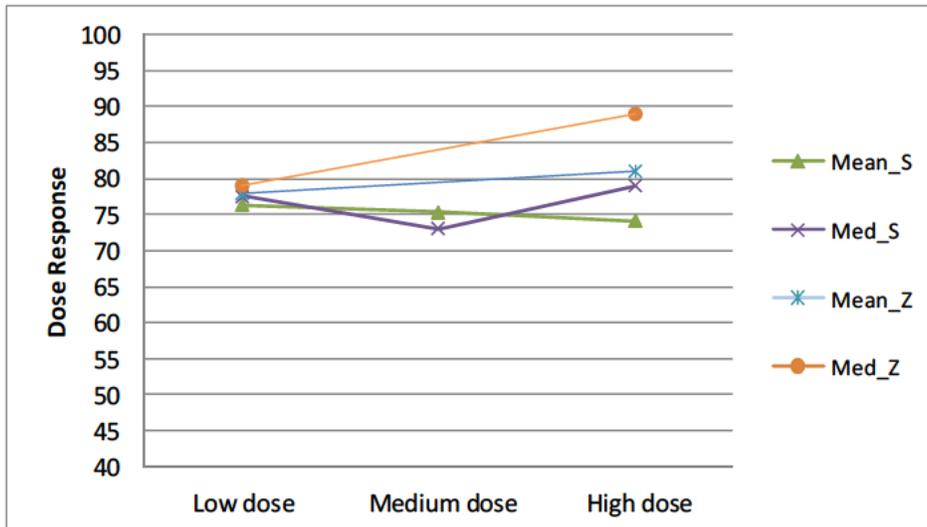
Table 1 shows that the third quartiles of S40, S150, Z15 and Z30 are 100. It means that even for a schedule IV drug, zolpidem, the Emax of Drug Liking VAS could reach the highest liking score in 25% of subjects or possibly more. One may notice that the means of the differences between suvorexant and zolpidem are all negative, and the medians of the differences are either zero or negative.

Figure 2 provides the distributions of five treatments as well as the differences between suvorexant and zolpidem by boxplots for Drug Liking VAS. The line in each box denotes the median and the circle in each box is for the mean. Because of the rules of the sorting process for characters in SAS, the boxplots related to S150 appear before those related to S40 and S80.

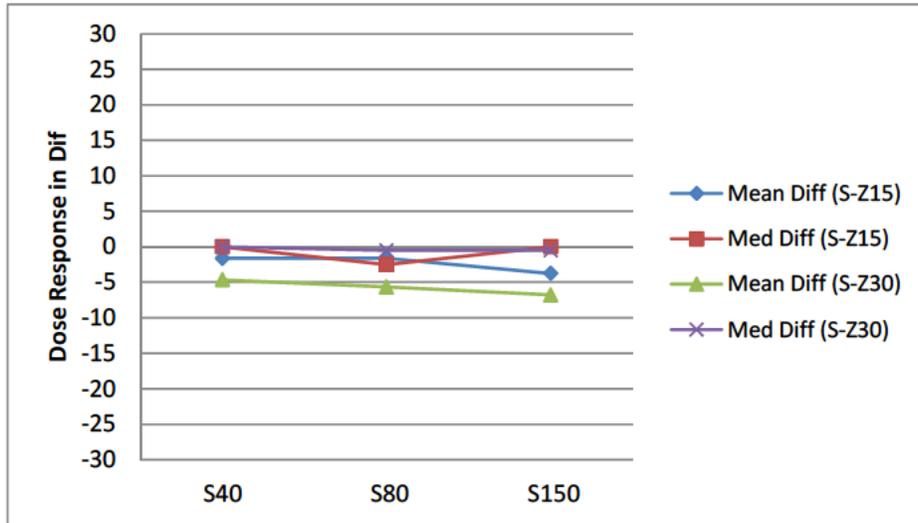


**Figure 2: Boxplots for five treatments and the differences between suvorexant and zolpidem for Drug Liking VAS (N=32)**

Figure 3 plots the mean and median dose response curves for suvorexant and zolpidem, and Figure 4 plots the mean and median dose response curves for suvorexant in difference from zolpidem for Drug Liking VAS. No apparent dose response of suvorexant is observed from either figure.

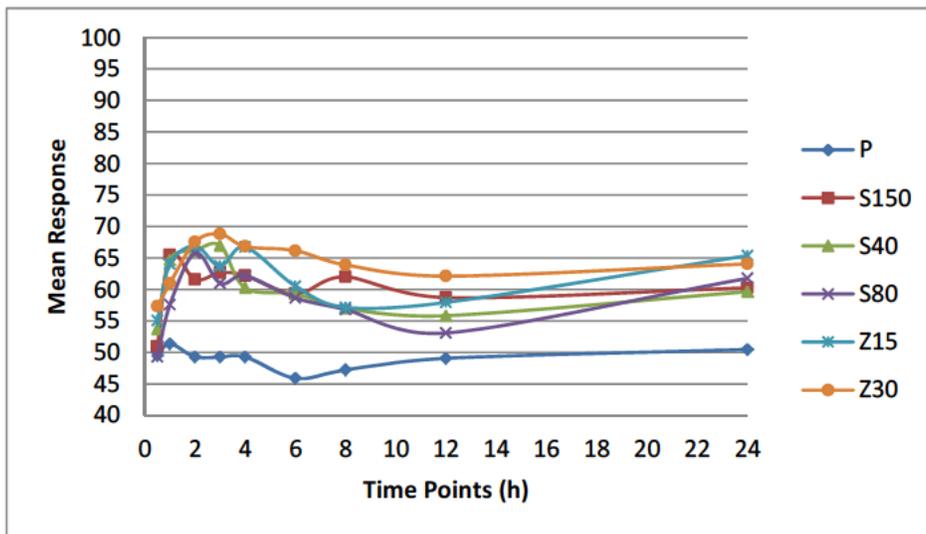


**Figure 3: Mean and median response curves for suvorexant and zolpidem for Drug Liking VAS**



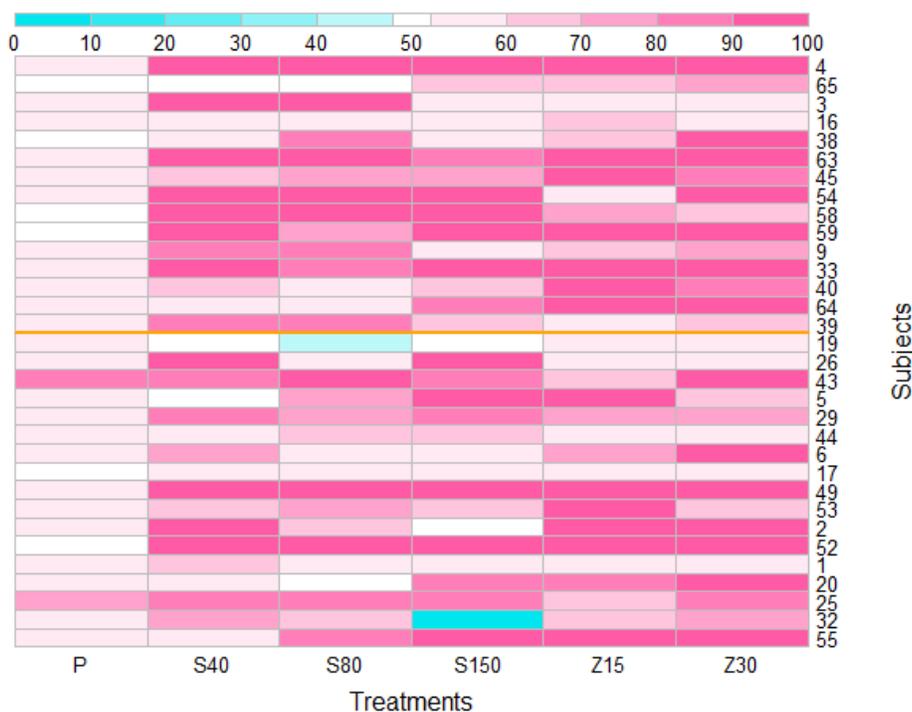
**Figure 4: Dose response curves of suvorexant in difference from zolpidem**

Figure 5 shows the mean time course profiles for Drug Liking VAS. If there are missing data at a time point, the mean is calculated using the existing data. With missing data in Z30, the mean time course profile of Z30 is still above those of other treatments. The differences in peak mean responses between suvorexant and zolpidem are less than 5 points. Three profiles from suvorexant lay between 57 and 67 except hour 0.5. The peak mean responses to S40, S80 and S150 are at hours 3, 2, and 1, respectively. The peak mean responses for Z15 and Z30 are at hour 2.



**Figure 5: Mean time course profiles for Drug Liking VAS**

Figure 6 is the heat map of the Emax of Drug Liking VAS by treatment.



**Figure 6: Heat map for Emax of Drug Liking VAS by treatment by subject**

As Figure 1, the blue, white and red in Figure 6 denote dislike, neutral and like, respectively. The orange line separates gender. This graph shows Emax of Drug Liking VAS from each subject for all treatments. Subjects #43 and #25 have high placebo responses. Most light pink responses for placebo are 51. From the heat map, overall there is no clear difference between three doses of suvorexant and two doses of zolpidem in Emax of Drug Liking VAS.

### 2.3.2.2 Statistical Testing

The statistical model used in the reviewer’s analysis was the mixed-effects model with period, sequence, treatment and gender as fixed effects, and subject as a random effect. Because no significant gender difference was found, the gender term was dropped from the model. The reviewer checked assumptions in the model for the equal variances and the normality. Neither assumption was satisfied. The SAS proc mixed procedure can adjust the unequal variances using Tukey-Kramer’s method. Thus, for the case that the residuals were not normally distributed, the reviewer checked the normality for the differences between treatments for each comparison. If the normality was satisfied, the paired t-test was used in the comparison otherwise the reviewer checked the assumption for the symmetry of the distribution of paired differences, if the distribution was approximately symmetric, Wilcoxon Signed-Rank test was used, otherwise the Sign-test was used. Table 2 lists the p-values for W-test, t-test, Wilcoxon Signed-Rank test, and Sign-test for the comparisons between suvorexant and zolpidem. The p-values in red are for the cases that W-test shows abnormality for the distribution of the paired differences. The green p-value indicates the test used for the comparison. For example, the p-value for the comparison between S40 and Z15 is 0.7141 which was based on the paired t-test, because the W-test was not significant.

**Table 2: P-values of the significance tests for Emax of Drug Liking VAS (n=32)**

Test	S40-Z15	S40-Z30	S80-Z15	S80-Z30	S150-Z15	S150-Z30
W	0.1775	0.0569	0.3436	0.3554	0.0257	0.0030
t	0.7141	0.2137	0.5364	0.1139	0.3587	0.0984
Signed- Rank	0.7539	0.2263	0.4420	0.1109	0.3415	0.0551
Sign	1.0000	0.3075	0.3449	0.3269	0.5413	0.0931

Note: All p-values were from two-sided tests.

The mean difference and the standard error of the mean difference as well as the median differences and the first and third quartiles of the differences for each comparison can be found in [Table 1](#). The test results show that there is no significant difference between each dose of suvorexant and any dose of zolpidem for Drug Liking VAS.

### 2.3.3 Secondary Analysis

The reviewer’s secondary analysis includes abuse potential measures: Bad Effects VAS, Bowdle VAS (External Perception and Bowdle Internal Perception), Good Effects VAS, High VAS, Overall Drug Liking VAS, and Take Drug Again VAS. The definitions of Bowdle External Perception and Bowdle Internal Perception can be found in [5.1 Appendix I](#).

The same methodologies as the primary analysis were used in the secondary analysis. Among all secondary measures, the normal assumption of the model is satisfied only for Bowdle External Perception and Bowdle Internal Perception. Table 3 summarizes the test results for Bowdle VAS. The adjusted p-values are from Tukey-Kramer’s method for unequal variances.

**Table 3: Summary of the test results for Emax of Bowdle VAS**

Measure	Treatment	S40	S80	S150	Z15	Z30	P
	N	32	32	32	32	32	32
Bowdle External Perception	LS mean	0.89	0.99	0.97	1.07	1.37	0.51
	95% CI	(0.70, 1.08)	(0.80, 1.19)	(0.77, 1.16)	(0.87, 1.26)	(1.17, 1.56)	(1.18, 1.56)
	Diff vs Z15	-0.27	-0.19	-0.22			
	Adj p-value	0.0443	0.3103	0.1796			
	Diff vs Z30	-0.49	-0.41	-0.43			
Adj p-value	<.0001	0.0003	<.0001				
Bowdle Internal Perception	LS mean	0.89	0.99	0.97	1.07	1.37	0.51
	95% CI	(0.69, 1.08)	(0.80, 1.19)	(0.77, 1.16)	(0.87, 1.26)	(1.17, 1.56)	(0.31, 0.70)
	Diff vs Z15	-0.18	-0.08	-0.1			
	Adj p-value	0.2917	0.9517	0.8497			
	Diff vs Z30	-0.48	-0.37	-0.4			
Adj p-value	<.0001	0.0004	0.0001				

Note: all p-values were from a two-sided test.

Table 3 shows that suvorexant has lower least square means than zolpidem for both Bowdle External Perception and Bowdle Internal Perception, and the differences are significant in the comparison between suvorexant and high dose of zolpidem, as well as in the comparison between suvorexant 40 mg and zolpidem 15 mg for Bowdle External Perception.

Tables 4-10 are summary statistics for five treatments in the study and for the treatment differences between suvorexant and zolpidem for Emax of Bad Effects VAS, Good Effects VAS, High VAS, Overall Drug Liking VAS, and Take Drug Again VAS as well as Bowdle VAS which includes Bowdle External Perception and Bowdle Internal Perception. Excluding Bowdle VAS, if the statistical test for the comparison between two treatments is significant based on the paired t test, the mean is highlighted in red in these tables; if the statistical test for the comparison between two treatments is significant based on the Sign test or the Wilcoxon Signed-Rank test, the median is highlighted in blue. The p-values of the tests are provided in [5.3 Appendix II](#).

**Table 4: Summary statistics and testing results for Emax of Bad Effects VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	17.63	5.07	0	0	0	49.5	99
S40	32	32.59	6.38	0	0	42.3	63	100
S80	32	45.88	5.93	0	11.5	58.0	72	100
S150	32	36.88	6.66	0	0.5	28.5	64.5	100
Z15	32	42.34	6.16	0	1.5	49.5	69.5	100
Z30	32	58.03	6.41	0	19.5	64.5	93.5	100
S40_Z15	32	-9.75	7.84	-100	-46.5	-1	13	100
S40_Z30	32	-25.44	7.98	-100	-57	-17.5	0	99
S80_Z15	32	3.53	7.25	-100	-23.5	0	37.5	79
S80_Z30	32	-12.16	8.11	-100	-47.5	-6.5	15	64
S150_Z15	32	-5.47	8.50	-100	-41.5	0	25.5	100
S150_Z30	32	-21.16	8.97	-100	-63.5	-12	4.5	87

**Table 5: Summary statistics and testing results for Emax of Bowdle External Perception**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	0.53	0.08	0.30	0.30	0.33	0.45	1.72
S40	32	0.86	0.11	0.30	0.32	0.56	1.63	1.97
S80	32	0.94	0.11	0.30	0.40	0.59	1.59	1.92
S150	32	0.91	0.10	0.30	0.37	0.74	1.51	2.00
Z15	32	1.13	0.09	0.30	0.70	0.99	1.72	2.00
Z30	32	1.35	0.10	0.33	0.77	1.69	1.78	1.99
S40_Z15	32	-0.27	0.09	-1.56	-0.50	-0.21	-0.03	0.85
S40_Z30	32	-0.48	0.10	-1.40	-1.05	-0.26	-0.03	0.21
S80_Z15	32	-0.20	0.10	-1.63	-0.50	-0.20	0.03	1.18
S80_Z30	32	-0.41	0.10	-1.46	-0.95	-0.25	-0.04	0.92
S150_Z15	32	-0.22	0.09	-1.63	-0.50	-0.19	0.05	1.10
S150_Z30	32	-0.44	0.11	-1.45	-1.00	-0.34	0.04	0.74

Note: The significant differences are highlighted in blue, and the test results are shown in Table 3.

**Table 6: Summary statistics and testing results for Emax of Bowdle Internal Perception**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	0.50	0.08	0.30	0.30	0.30	0.39	1.72
S40	32	0.90	0.10	0.30	0.40	0.66	1.53	2.00
S80	32	0.99	0.10	0.30	0.49	0.84	1.72	1.91
S150	32	0.97	0.10	0.30	0.55	0.72	1.48	2.01
Z15	32	1.07	0.09	0.30	0.63	0.90	1.66	2.01
Z30	32	1.36	0.09	0.34	0.93	1.63	1.74	2.01
S40_Z15	32	-0.18	0.07	-1.24	-0.34	-0.03	0.04	0.48
S40_Z30	32	-0.47	0.08	-1.46	-0.76	-0.29	-0.06	0.16
S80_Z15	32	-0.08	0.10	-1.22	-0.47	-0.07	0.14	1.18
S80_Z30	32	-0.38	0.09	-1.39	-0.74	-0.32	-0.04	1.24
S150_Z15	32	-0.11	0.08	-1.23	-0.30	-0.05	0.13	1.13
S150_Z30	32	-0.40	0.11	-1.45	-0.87	-0.25	0.02	0.84

Note: The significant differences are highlighted in blue, and the test results are shown in Table 3.

**Table 7: Summary statistics and testing results for Emax of Good Effects VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	21.41	4.86	0	0	0	50	80
S40	32	74.84	5.43	0	56.5	87	100	100
S80	32	75.59	4.01	0	66	73.5	98.5	100
S150	32	74.53	5.18	0	51	84	100	100
Z15	32	82.56	3.51	26	69.5	87	100	100
Z30	32	84.97	2.75	51	73.5	88	100	100
S40_Z15	32	-7.72	6.43	-100	-22	0	17	49
S40_Z30	32	-10.13	5.37	-99	-19.5	0	2	40
S80_Z15	32	-6.97	4.36	-83	-19	0	8	34
S80_Z30	32	-9.38	4.25	-100	-25	-1.5	3	28
S150_Z15	32	-8.03	4.71	-83	-28	0	8	34
S150_Z30	32	-10.44	5.40	-100	-27.5	0	2	49

**Table 8: Summary statistics and testing results for Emax of High VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	17.53	4.47	0	0	0	47	85
S40	32	67.69	5.91	0	58.5	70.5	97.5	100
S80	32	65.34	6.05	0	50.5	72	98.5	100
S150	32	71.03	5.55	0	54.5	82	96	100
Z15	32	71.19	5.51	0	58.5	78	99	100
Z30	32	83.94	3.20	42	69.5	88.5	100	100
S40_Z15	32	-3.50	7.16	-100	-30	0	19.5	100
S40_Z30	32	-16.25	4.98	-99	-32.5	-5	1	36
S80_Z15	32	-5.84	6.74	-100	-25.5	-1	13	99
S80_Z30	32	-18.59	5.42	-100	-42.5	-6.5	0	24
S150_Z15	32	-0.16	6.35	-80	-23	0.5	25	91
S150_Z30	32	-12.91	5.36	-96	-28.5	-8	1	45

**Table 9: Summary statistics and testing results for Emax of Overall Drug Liking VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	48.56	2.30	0	50	50	51	84
S40	32	70.09	4.69	0	50.5	75.5	96	100
S80	32	69.97	4.56	3	54	74	91.5	100
S150	32	67.84	5.25	0	50	77.5	92.5	100
Z15	32	71.75	3.99	25	51	76.5	94	100
Z30	32	73.38	4.30	5	53	80.5	100	100
S40_Z15	32	-1.66	5.69	-84	-13	0	11.5	61
S40_Z30	32	-3.28	6.38	-100	-10.5	0	12	92
S80_Z15	32	-1.78	5.32	-81	-18	0	16	63
S80_Z30	32	-3.41	5.47	-97	-17.5	0	7.5	52
S150_Z15	32	-3.91	5.17	-65	-20.5	-2	11.5	61
S150_Z30	32	-5.53	5.29	-59	-16	-3.5	0.5	95

**Table 10: Summary statistics and testing results for Emax of Take Drug Again VAS**

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
P	32	35.53	5.31	0	0.5	50	51	100
S40	32	71.78	5.27	0	50.5	77.5	100	100
S80	32	68.84	5.40	0	50	73.5	100	100
S150	32	75.56	5.07	0	60.5	84.5	100	100
Z15	32	72.53	5.21	0	53.5	77	100	100
Z30	32	74.75	5.35	2	59	86.5	100	100
S40_Z15	32	-0.75	5.78	-75	-11.5	0	3	79
S40_Z30	32	-2.97	7.05	-100	-14.5	0	14.5	98
S80_Z15	32	-3.69	5.95	-77	-25.5	0	8.5	73
S80_Z30	32	-5.91	6.92	-98	-30	0	7	72
S150_Z15	32	3.03	7.13	-100	-2.5	0	8.5	100
S150_Z30	32	0.81	5.90	-51	-11	0	9	98

There are 36 comparisons in the above 6 tables excluding Tables for Bowdle External Perception and Bowdle Internal Perception, and Bad Effects VAS. The mean differences for suvorexant are lower than those for zolpidem in 34 comparisons, and the median differences between suvorexant and zolpidem are less than or equal to 0 except S150 versus Z15 for High VAS (the median difference is 0.5). Excluding Bowdle External Perception and Bowdle Internal Perception (the significant results can be found in [Table 3](#)), the significance results are found for the comparisons S40 versus Z30 and S150 versus Z30 for Bad Effects VAS, S80 versus Z30 and S150 versus Z30

for High VAS, and S150 versus Z30 for Overall Drug Liking VAS in the reviewer’s secondary analysis.

Figures 7-13 are the mean time course profiles by treatment for these measures [5.2 Appendix II](#).

### 3. Conclusion

The results from the reviewer’s analyses are summarized in Table 11.

**Table 11: Summary of the results from significance tests for the abuse potential measures considered in this review**

Study P025	Abuse Potential Measure	S40 vs. Z15	S80 vs. Z15	S150 vs. Z15	S40 vs. Z30	S80 vs. Z30	S150 vs. Z30
Primary	Drug Liking VAS	NS	NS	NS	NS	NS	NS
Secondary	Bad Effects VAS	NS	NS	NS	S (<)	NS	S (<)
	Bowdle External Perception	S (<)	NS	NS	S (<)	S (<)	S (<)
	Bowdle Internal Perception	NS	NS	NS	S (<)	S (<)	S (<)
	Good Effects VAS	NS	NS	NS	NS	NS	NS
	High VAS	NS	NS	NS	NS	S (<)	S (<)
	Overall Drug Liking VAS	NS	NS	NS	NS	NS	S (<)
	Take Drug Again VAS	NS	NS	NS	NS	NS	NS

This study shows that

- There is no significant mean (or median) difference between suvorexant and zolpidem for Drug Liking VAS, Good Effects VAS, and Take Drug Again.
- On the average, the responses to three doses of suvorexant are significantly lower than those to zolpidem 30 mg for Bowdle External Perception and Bowdle Internal Perception. For Bowdle External Perception, the responses to suvorexant 40 mg is also significantly lower than those to zolpidem 15mg.
- On the average, the responses to suvorexant 150 mg are significantly lower than those of zolpidem 30 mg for Bad Effect VAS, High VAS, and Overall Drug Liking VAS.
- On the average, the responses to suvorexant 40 mg and 80 mg are significantly lower than those to zolpidem 30 mg for Bad Effects VAS and High VAS respectively.
- The abuse potential of suvorexant is greater than that of placebo.
- The study is validated by the comparison between two doses of zolpidem and placebo for Drug Liking VAS.
- There is no apparent dose-response for suvorexant.

In summary, suvorexant has higher abuse liability than placebo, and does not demonstrate lower abuse liability than zolpidem for the primary measure, Drug Liking VAS. Considering the results from both the reviewer’s primary and secondary analyses, and the fact that there is no apparent

dose-response for suvorexant, the reviewer concludes that the abuse potential of suvorexant may be similar to or lower than zolpidem.

#### 4. References

Chen , L, Bonson, RK. An Equivalence Test for the Comparison between a Test Drug and Placebo in Human Abuse Potential Studies. *Journal of Biopharmaceutical Statistics*. 2013; 23: 294-306.

Bowdle, TA, Radant, AD, Cowley, DS, Kharasch, ED, Strassman, RJ, Roy-Byrne, PP. Psychedelic effects of ketamine in healthy volunteers: relationship to steady-state plasma concentrations. *Anesthesiology*. 1998; 88: 82–88.

Chen, L, Wang, Y. Heat Map Displays for Data From Human Drug Abuse Potential Crossover Studies. *Drug Information Journal*. (2012) 46:6, 701-707.

## 5. Appendices

### 5.1 Appendix I

The Bowdle VAS [Bowdle, *et al* 1998] consists of 13 items for which the subject is asked to rate his/her current feelings. Each VAS will be scored from 0 to 100, with 0 reflecting “not at all” and 100 reflecting “extremely”. Lower individual and overall scores indicate fewer psychedelic effects.

The individual items of the questionnaire are listed below:

1. My body or body parts seemed to change their shape or position\*
2. My surroundings seemed to change in size, depth, or shape\*
3. The passing of time was altered\*
4. I had feelings of unreality\*\*
5. It was difficult to control my thoughts\*
6. The intensity of colors changed\*
7. The intensity of sound changes\*
8. I heard voices or sounds that were not real\*\*
9. I had the idea that events, objects, or other people had particular meaning that was specific for me\*\*
10. I had suspicious ideas or the belief that others were against me\*\*
11. I felt anxious\*\*
12. I felt high 1
13. I felt drowsy 1

1 Items 12 and 13 was not included, as they are part of the individual VAS items

\* Items 1, 2, 3, 5, 6, and 7 was combined to assess the derived variable “subjective external perception”.

\*\* Items 4, 8, 9, 10, and 11 was combined to assess the derived variable “subjective internal perception”

**External perception ( $\log_{10}(mm)$ ):** average of  $\log(2+Item1)$ ,  $\log(2+Item2)$ ,  $\log(2+Item3)$ ,  $\log(2+Item5)$ ,  $\log(2+Item6)$ ,  $\log(2+Item7)$

**Internal perception ( $\log_{10}(mm)$ ):** average of  $\log(2+Item4)$ ,  $\log(2+Item8)$ ,  $\log(2+Item9)$ ,  $\log(2+Item10)$ ,  $\log(2+Item11)$

In the case that one of the items was missing the related score was not calculated.

## 5.2 Appendix II

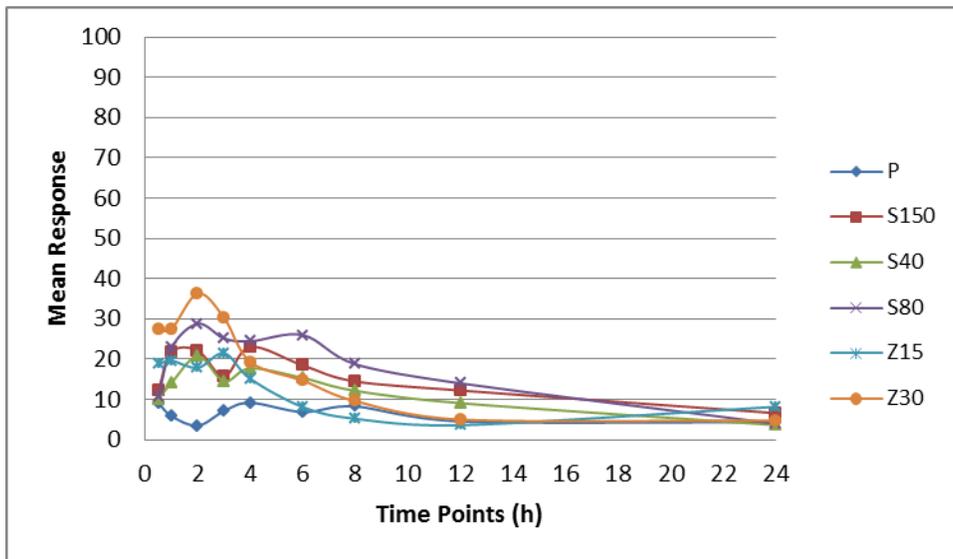


Figure 7: Mean time course profiles for Bad Effects VAS

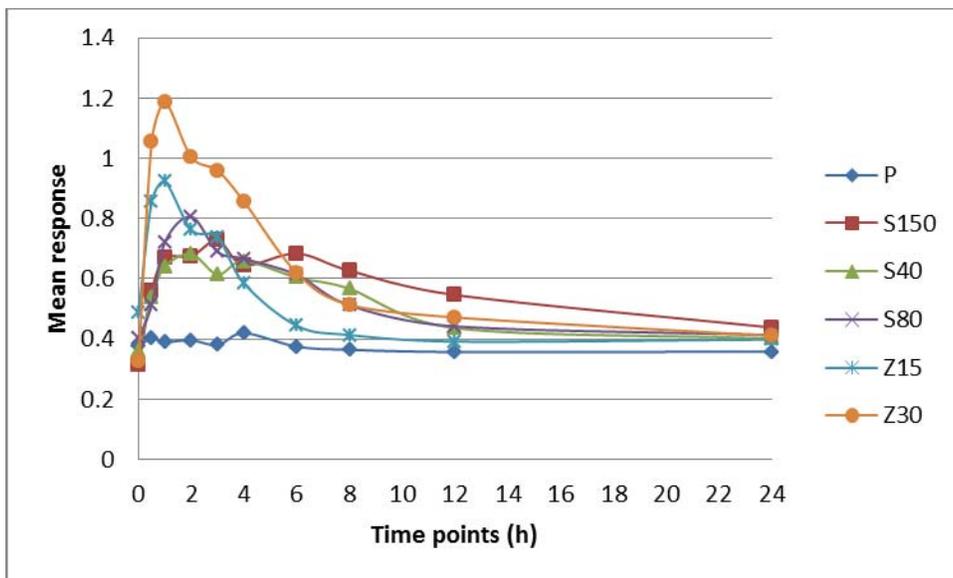


Figure 8: Mean time course profiles for Bowdle External Perception

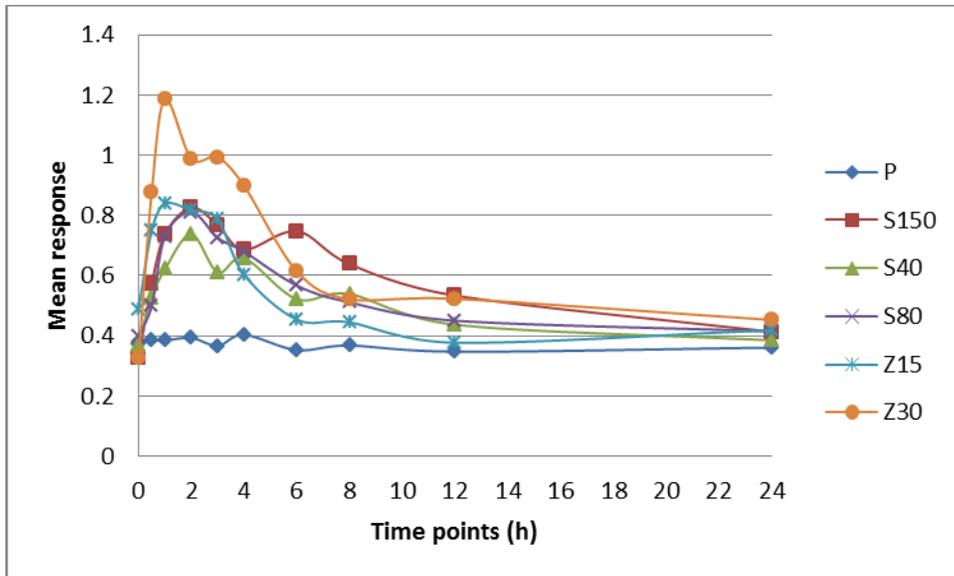


Figure 9: Mean time course profiles for Bowdle Internal Perception

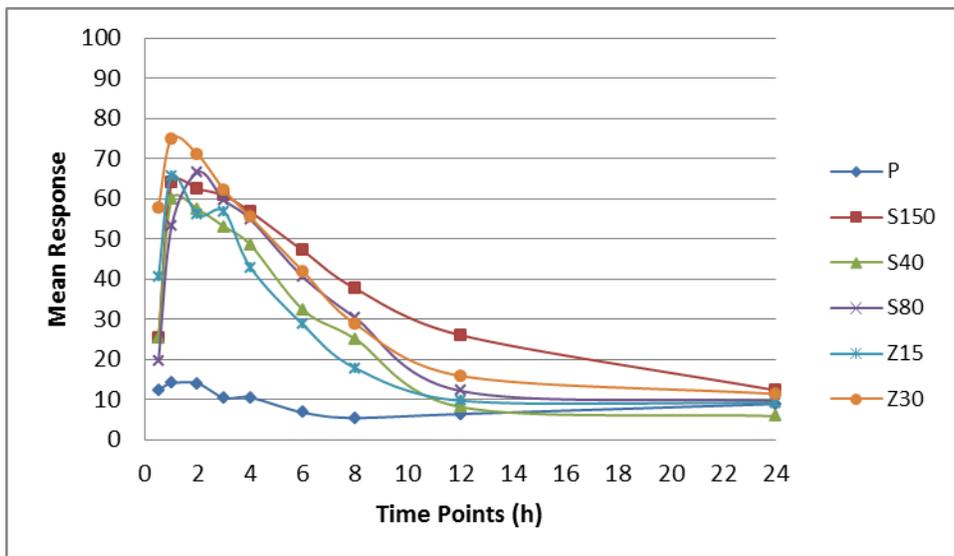


Figure 10: Mean time course profiles for Good Effects VAS

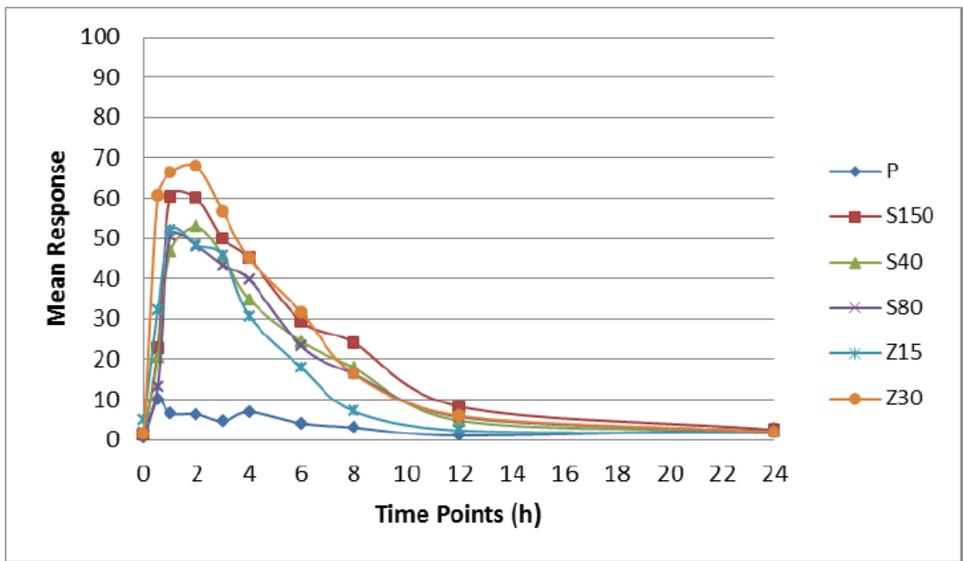


Figure 11: Mean time course profiles for High VAS

N

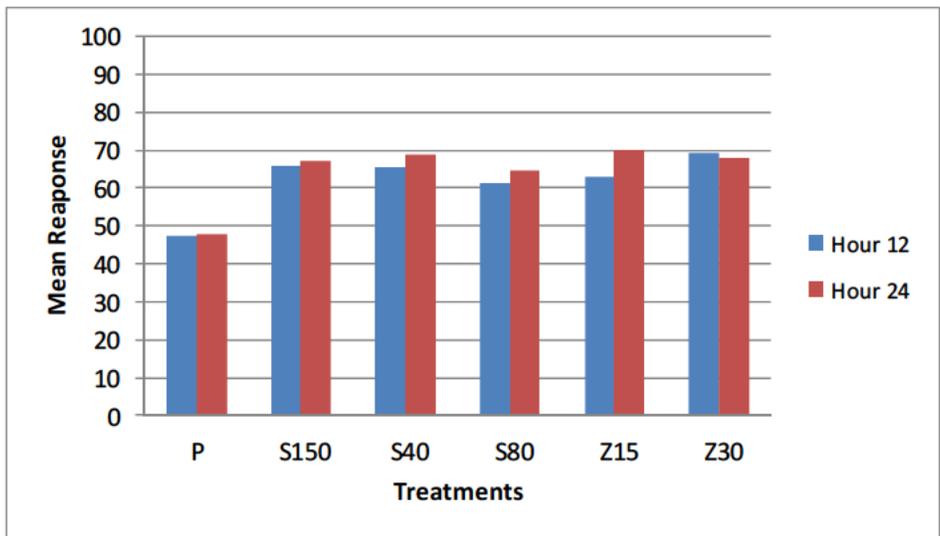
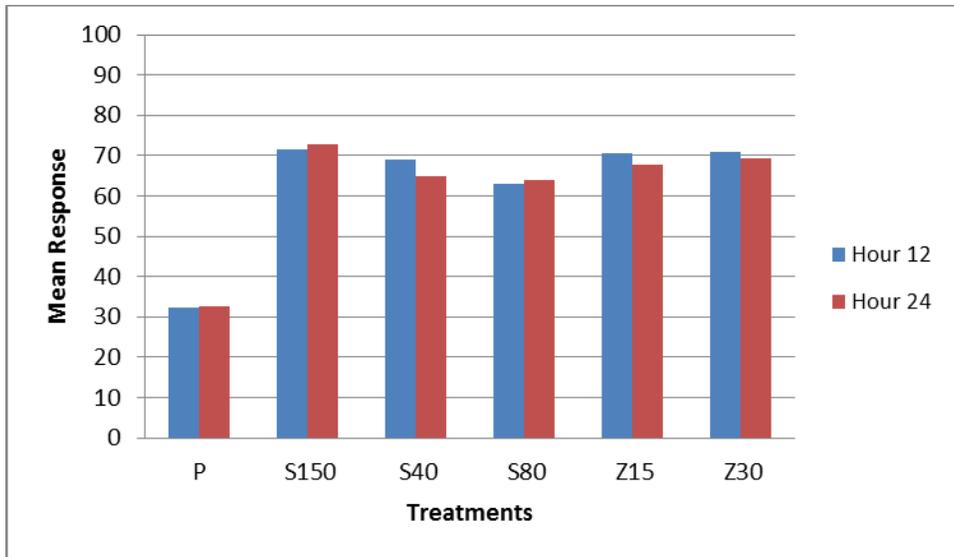


Figure 12: Overall Drug Liking VAS at hours 12 and 24



**Figure 13: Take Drug Again VAS at hours 12 and 24**

### 5.3 Appendix III

**Table 12: P-values of the significance tests for Emax of secondary measures excluding Bowdle External Perception and Bowdle Internal Perception (n=32)**

Measure	Test	S40-Z15	S40-Z30	S80-Z15	S80-Z30	S150-Z15	S150-Z30
Bad effects VAS	t	0.2231	0.0033	0.6298	0.1439	0.5246	0.0248
	Sign	0.3449	0.0009	0.7011	0.2153	1.0000	0.0987
	SignRank	0.2077	0.0005	0.5037	0.2085	0.7477	0.0281
	W	0.3237	0.0322	0.5157	0.1213	0.1163	0.1569
Good Effects VAS	t	0.2394	0.0687	0.1205	0.0349	0.0979	0.0622
	Sign	0.4244	0.6900	0.7011	0.3449	0.5235	0.7011
	SignRank	0.6472	0.1702	0.2131	0.0453	0.1479	0.1290
	W	0.0034	0.0039	0.0654	0.0009	0.0085	0.0045
High VAS	t	0.6285	0.0027	0.3924	0.0017	0.9805	0.0222
	Sign	1.0000	0.1078	0.5847	0.0125	0.7111	0.0522
	SignRank	0.6886	0.0022	0.4559	0.0013	0.9329	0.0173
	W	0.2361	0.0408	0.3416	0.0027	0.8819	0.1594
Overall Drug Liking VAS	t	0.7731	0.6107	0.7402	0.5380	0.4556	0.3036
	Sign	1.0000	1.0000	0.8506	0.8388	0.2478	0.0241
	SignRank	0.9070	0.9312	0.7477	0.6273	0.3584	0.0483
	W	0.0777	0.0073	0.4416	0.0514	0.5340	0.0023
Take Drug Again VAS	t	0.8976	0.6767	0.5400	0.3998	0.6737	0.8914
	Sign	0.8318	1.0000	1.0000	0.2100	0.6776	1.0000
	SignRank	0.7298	0.8355	0.5547	0.4237	0.5951	0.8463
	W	0.0042	0.0319	0.1767	0.1430	0.0018	0.0006

Note: The normality of the model is not satisfied for the measures in this table. The red p-value indicates that the W-test is significant for the distribution of the paired differences, and the green p-value indicates that the test used in the evaluation is according to the assumption of the test. For example, In the case of S40-Z15 for High VAS, the W test is not significant. It means that the normal assumption for the distribution of differences in responses between S40 and Z15 is not violated. Thus, the t-test is used for the comparison, and resulted in a p-value of 0.6285 (two-sided).

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

LING CHEN  
05/01/2013

STELLA G MACHADO  
05/01/2013

# Addendum-1

[Subject: Corrected Table for Rat Carcinogenicity Data Analysis]



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Science  
Office of Biostatistics

**IND/NDA Number:** NDA 204-569

**Drug Name:** Suvorexant

**Indication(s):** 104 Week Rat and 26 Week Mouse Carcinogenicity Studies

**Applicant:** Sponsor: Merck Sharp & Dohme Corp.  
P.O. Box 2000, RY33-208,  
Rahway NJ 07065

**Documents Reviewed:** Electronic submission submitted on August 30, 2012  
Electronic data submitted on August 30, 2012

**Review Priority:** Standard

**Biometrics Division:** Division of Biometrics -6

**Statistical Reviewer:** Mohammad Atiar Rahman, Ph.D.

**Concurring Reviewer:** Karl Lin, Ph.D.

**Medical Division:** Division of Neurological Products

**Reviewing Pharmacologist:** Richard Siarey, Ph.D.

**Project Manager:** Cathleen B. Michaloski

**Keywords:** Carcinogenicity, Dose response

**Introduction:** A statistical review of this submission was issued by this reviewer on January 10, 2013. In a later read-through this reviewer found some errors in his calculations in the rat data analysis. This addendum contains the corrected results.

**New tables:** Tables 3A and 3B on Pages 4 through 7.

**Effects of new results:** The re-calculation changed the p-values a little; however, the conclusions remained the same.

Mohammad Atiar Rahman, Ph.D.  
Mathematical Statistician

Concur: Karl Lin, Ph.D.  
Team Leader, Biometrics-6

cc:  
Archival NDA 204-569  
Dr. Siarey  
Ms. Michaloski

Dr. Machado  
Dr. Lin  
Dr. Rahman  
MS. Patrician







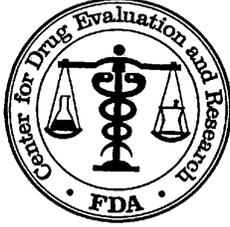


-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

MOHAMMAD A RAHMAN  
01/24/2013

KARL K LIN  
01/24/2013  
Concur with review



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Science  
Office of Biostatistics

## Statistical Review and Evaluation

### CARCINOGENICITY STUDIES

**IND/NDA Number:** NDA 204-569

**Drug Name:** Suvorexant

**Indication(s):** 104 Week Rat and 26 Week Mouse Carcinogenicity Studies

**Applicant:** Sponsor: Merck Sharp & Dohme Corp.  
P.O. Box 2000, RY33-208,  
Rahway NJ 07065

**Documents Reviewed:** Electronic submission submitted on August 30, 2012  
Electronic data submitted on August 30, 2012

**Review Priority:** Standard

**Biometrics Division:** Division of Biometrics -6

**Statistical Reviewer:** Mohammad Atiar Rahman, Ph.D.

**Concurring Reviewer:** Karl Lin, Ph.D.

**Medical Division:** Division of Neurological Products

**Reviewing Pharmacologist:** Richard Siarey, Ph.D.

**Project Manager:** Cathleen B. Michaloski

**Keywords:** Carcinogenicity, Dose response

**Table of Contents**

1. .... Background 3

2. .... Rat Study 3

    2.1. Sponsor's analyses.....3

        2.1.1. Survival analysis.....3

**Sponsor's findings** .....3

        2.1.2. Tumor data analysis.....3

**Sponsor's findings** .....4

    2.2. Reviewer's analyses .....4

        2.2.1. Survival analysis.....4

**Reviewer's findings** .....4

        2.2.2. Tumor data analysis.....4

**Multiple testing adjustment** .....5

**Reviewer's findings** .....5

3. .... Mouse Study 6

    3.1. Sponsor's analyses.....6

        3.1.1. Survival analysis.....6

        3.1.2. Tumor data analysis.....7

**Sponsor's findings** .....7

    3.2. Reviewer's analyses .....7

        3.2.1. Survival analysis.....7

**Reviewer's findings** .....7

        3.2.2. Tumor data analysis.....8

**Reviewer's findings** .....8

4. .... Summary 8

5. .... Appendix 11

References ..... 25

## 1. Background

In this submission the sponsor included reports of two animal carcinogenicity studies, one in regular rats and one in transgenic mice. These studies were intended to assess the carcinogenic potential of MK4305 (Suvorexant) when administered orally by gavage once daily at appropriate drug levels for 104 weeks in rats and for 26 weeks in mice. Results of this review have been discussed with the reviewing pharmacologist Dr. Siarey.

In this review the phrase "dose response relationship" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

## 2. Rat Study

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were three treated groups and two identical vehicle control groups. Two hundred and fifty Sprague -Dawley, CrI:CD(SD) rats of each sex were randomly allocated to treated and control groups in equal size of 50 rats. The dose levels for treated groups were 80, 160, or 325 mg/kg/day for males and 40, 80, or 325 mg/kg/day for females. In this review these dose groups would be referred to as the low, medium, and high dose groups, respectively. Control 1 and Control 2 rats were given 1300 mg/kg/day of HPMC (hydroxypropyl methylcellulose-acetate succinate) in 0.5% methylcellulose with 5 mM hydrochloric acid in deionized water (vehicle) by gavage.

During the administration period all rats were observed daily for mortality and morbidity. All rats were observed weekly for clinical signs through Study Week 104. Beginning in Study Week 26, all rats were palpated for masses every 4 weeks through Study Week 104. Additionally, unscheduled palpations were performed on rats found with masses in Study Weeks 6, 10, 14, 18, 23, 39, 44, 52, 53, 59, 60, 61, 64, 68, and generally weekly thereafter.

Body weights for all rats were measured at pretest, once per week from Study Week 1 through Study Week 14, and once every 4 weeks thereafter through Study Week 102.

### 2.1. Sponsor's analyses

#### 2.1.1. Survival analysis

Survival function of each treatment group was estimated using the Kaplan-Meier product limit method and was presented graphically. An overall test comparing all groups was conducted using a log-rank test.

**Sponsor's findings:** The sponsor's analysis showed 40%, 52%, 48%, 28%, and 42% mortality of male rats and 32%, 46%, 34%, 28%, and 30% mortality of female rats in control 1, control 2, low, medium, and high dose groups, respectively. The sponsor concluded that compared to control groups there were no study drug related effects on mortality rates. The sponsor commented that the causes of death were of the types seen in the untreated rats in this laboratory.

#### 2.1.2. Tumor data analysis

Tumor incidence data were analyzed using the methods outlined in the paper of Peto et al. (1980). The pair wise comparisons of control groups with the treated groups were conducted using the Fisher's exact test.

**Adjustment for multiple testing:** The sponsor adjusted the effects of multiple testing using the SAS Proc MultTest for Peto's test. The SAS Proc MultTest generally follows the recommendations made by Mantel, Heyse and Rom, and Harter.

**Sponsor's findings:** Sponsor's analyses showed a statistically significant increased incidence of hepatocellular adenomas in male rat high dose group, and thyroid follicular cell adenomas in male rat medium and high dose group and in female rat high dose group.

## 2.2. Reviewer's analyses

To verify sponsor's analyses and to perform additional analysis suggested by the reviewing pharmacologist, this reviewer independently performed survival and tumor data analyses. Data used in this reviewer's analyses were provided by the sponsor electronically.

The submitted rat study has two identical vehicle control groups. For studies with two identical controls, the FDA statistical guidance for carcinogenicity data analysis suggests to analyze the data combining the two control groups. Following the guidance suggestion, this reviewer analyzed both the mortality and tumor data using the combined control.

### 2.2.1. Survival analysis

The survival distributions of rats in all five treatment groups were estimated by the Kaplan-Meier product limit method. For combined control, low, medium, and high dose groups, the dose response relationship was tested using the likelihood ratio test and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1A and 1B in the appendix for male and female rats, respectively. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for male and female rats, respectively. Results of the tests for dose response relationship and homogeneity of survivals, are given in Tables 2A and 2B in the appendix for male and female rats, respectively.

**Reviewer's findings:** This reviewer's analysis showed 40%, 52%, 48%, 28%, and 42% mortality of male rats and 32%, 46%, 34%, 28%, and 28% mortality of female rats in control 1, control 2, low, medium, and high dose groups, respectively. Tests did not show statistically significant dose response relationship in mortality across combined control and treated groups in either sex. The pairwise comparisons showed statistically significant decreased mortality in the male rat medium dose group compared to combined control.

**Reviewer's comment:** *The sponsor's calculation showed 30% death in the female high dose group, and this reviewer's calculation showed 28%. This difference is due to the fact that one female rat (# 090539) from high dose group died due to natural causes during the terminal sacrifice weeks. The sponsor counted this with the natural deaths, while this reviewer counted this with the terminally sacrificed rats.*

### 2.2.2. Tumor data analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups. Both the dose response relationship tests and pairwise comparisons were performed using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993). In this method an animal that lives the full study period ( $w_{\max}$ ) or dies before the terminal sacrifice but develops the tumor type being tested gets a score of  $s_h = 1$ . An animal that dies at week  $w_h$  without a tumor before the end of

the study gets a score of  $s_h = \left( \frac{w_h}{w_{max}} \right)^k < 1$ . The adjusted group size is defined as  $\sum s_h$ . As an interpretation, an

animal with score  $s_h = 1$  can be considered as a whole animal while an animal with score  $s_h < 1$  can be considered as a partial animal. The adjusted group size  $\sum s_h$  is equal to N (the original group size) if all animals live up to the end of the study or if each animal that dies before the terminal sacrifice develops at least one tumor, otherwise the adjusted group size is less than N. These adjusted group sizes are then used for the dose response relationship (or the pairwise) tests using the Cochran-Armitage test. One critical point for Poly-k test is the choice of the appropriate value of k, which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. Hence, this reviewer used k=3 for the analysis of this data. For the calculation of p-values the exact permutation method was used. The tumor rates and the p-values of the tested tumor types are listed in Tables 3A and 3B in the appendix for male and female rats, respectively.

**Multiple testing adjustment:** Noting that present submission had one long term study in rats and one medium term study in mouse, the adjustment for multiple testing of dose response relationship was conducted using the division of biometrics recommendation i.e. for dose response relationship tests use test levels of  $\alpha=0.005$  for common tumors and  $\alpha=0.025$  for rare tumors in rat study and use test levels of  $\alpha=0.05$  for all tumors in mouse study. For pairwise comparisons of treated group with control use levels  $\alpha=0.01$  for common tumors and  $\alpha=0.05$  for rare tumors in rat study and use  $\alpha=0.05$  for all tumor types in mouse study.

**Reviewer’s findings:** The following tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of combined control and treated groups.

**Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups With Combined Control in Rats**

Organ Name	Tumor Name	Com C N=100	Low N=50	Med N=50	Hi gh N=50	P_Val ue			
						Dose Resp	Com C vs. L	Com C vs. M	Com C vs. H
<b>Male rats</b>									
Brain	Granular cell tumor	0	0	0	2	0.0388	.	.	0.1169
Liver	Hepatocel lular adenoma	1	0	2	7	<0.001*	1.0000	0.3178	0.0024*
	Hepatocel lular adn+car	1	0	2	9	<0.001*	1.0000	0.3178	<0.001*
	Hepatocel lular carci noma	0	0	0	2	0.0404	.	.	0.1208
Skin	Hemangi osarcoma	0	0	0	2	0.0404	.	.	0.1208
Thyroi d/Fol l i cu	Adenoma	0	2	4	13	<0.001*	0.1130	0.0196*	<0.001*
	Adn+Car	0	3	4	13	<0.001*	0.0370*	0.0196*	<0.001*
<b>Female rats</b>									
Thyroi d/Fol l i cu	Adenoma	2	0	0	10	<0.001*	1.0000	1.0000	<0.001*
	Adn+Car	2	0	1	11	<0.001*	1.0000	0.7279	<0.001*

\* Com C: Combined vehicle control

Based on the criteria of adjustment for multiple testing discussed above, the incidence of hepatocellular

adenoma, and combined incidences of hepatocellular adenoma and carcinoma in male rats, and incidence of thyroid follicular cell adenoma, combined incidences of thyroid follicular cell adenoma and carcinoma in both sexes were considered to have statistically significant dose response relationships. The pairwise comparisons showed statistically significant increased incidence of hepatocellular adenoma, combined incidences of hepatocellular adenoma and carcinoma in the male rat high dose group compared to the combined control. The pairwise comparisons also showed statistically significant increased incidence of thyroid follicular cell adenoma, and combined incidences of thyroid follicular cell adenoma and carcinoma in male rat medium and high dose groups, and in female rat high dose group compared to the combined control.

### 3. Mouse Study

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were four treated groups, two vehicle control groups, and one positive control group. One hundred and seventy five male and female ic:CB6F1-TgrasH2@Jcl mice were assigned randomly to the treated, vehicle control and positive control groups in equal size of 25<sup>#</sup> mice per group. The dose levels for treated groups were 25, 50, 200, and 650 mg/kg/day for both sexes. In this review these dose groups were referred to as the low, medium, mid-hi, and high dose group, respectively. The vehicle controls received the vehicle (sterile water for injection) by gavage. The positive control mice were dosed via intraperitoneal (i.p.) injections of urethane (L-000471271) in saline on Days 1, 3 and 5, at a dose level of 1000 mg/kg/day. A dose volume of 10 mL/kg body weight was used for all groups.

All mice except those from the positive control group were observed daily for mortality, and once a week for physical signs through Study Week 26. All mice from the positive control group were observed daily for mortality, and on the dosing days, and then once per week from Study Week 2 to Study Week 26 for physical signs. All mice except those from the positive control group were palpated for masses every 4 weeks, from Study Week 14 to Study Week 26. Body Weights of all mice were measured pretest and once per week from Study Week 1 through Study Week 26.

#### 3.1. Sponsor's analyses

##### 3.1.1. Survival analysis

The sponsor presented a summary table of the mortalities of mice by sex. Survival function of each treatment group was estimated using the Kaplan-Meier product limit method and was presented graphically. An overall test of homogeneity/equality of survival function was conducted using the log rank test based on the censored data generalization of the Savage (exponential score) test. These tests were performed using the SAS Proc Life Test.

**Sponsor's findings:** The sponsor analysis showed 0, 4, 24, 0, 0, 0, and 1 deaths of male mice and 1, 0, 24, 1, 1, 1, and 2 deaths of female mice in vehicle control 1, vehicle control 2, positive control, low, medium, mid-hi, and high dose groups, respectively. The sponsor's analysis did not show statistically significant positive dose response relationship in mortality in either sex. The analysis however showed a statistically significant decrease ( $p=0.037$ ) in mortality through the dose of 200 mg/kg/day. The positive control showed statistically significant increased mortality compared to the controls.

---

<sup>#</sup> The sponsor mentioned that there were 27 mice in each treatment groups, which included 25 study mice and 2 extra mice. The submitted data sets had data from 25 mice per group.

The sponsor concluded that there were no unscheduled deaths related to the administration of the test article. The increase in unscheduled deaths was mainly due to neoplastic processes such as hemangiosarcoma in the spleen and adenocarcinoma in the lung. The causes of death or the reasons for early sacrifice in the treated groups were those expected for this species.

### 3.1.2. Tumor data analysis

The sponsor analyzed the tumor data using the method proposed by Peto et al. (1980) for dose response relationships on combined vehicle control, and four treated groups. The pairwise comparisons of treated groups and positive control group with combined control group were performed using the Fisher exact test.

**Adjustment for multiple testing:** No adjustment for multiple testing was performed.

**Sponsor's findings:** The sponsor's analysis did not show statistically significant increased incidence in any of the observed tumor types in the treated groups in either sex. In positive control, there were statistically significant increases in incidences of the following tumors: adenoma and adenocarcinoma in the lung in both sexes, hemangiosarcoma in the spleen in both sexes and in the lung in females.

The sponsor concluded that there was no test article-related evidence of carcinogenic potential up to 650 mg/kg/day in the study mice in either sex.

## 3.2. Reviewer's analyses

Similar to the rat study, to verify sponsor's analyses and to perform additional analysis suggested by the reviewing pharmacologist, this reviewer independently performed survival and tumor data analyses. Data used in this reviewer's analyses were provided by the sponsor electronically.

For the analysis of both the survival data and the tumor data this reviewer used similar methods as he used for the analysis of the rat data.

Similar to rat study, the submitted mouse study also has two identical vehicle control groups. Following the FDA carcinogenicity data analysis guidance, this reviewer combined the two vehicle controls and analyzed both the mortality and tumor data using the combined control.

### 3.2.1. Survival analysis

The intercurrent mortality data of all treatment groups are given in Tables 4A and 4B in the appendix for male and female mice, respectively. Results of the tests for dose response relationship and homogeneity of survivals for combined control, low, medium, mid-hi, and high dose groups are given in Tables 5A and 5B in the appendix for male and female mice, respectively. The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 2A and 2B in the appendix for male and female mice, respectively.

**Reviewer's findings:** This reviewer's analysis showed 0, 4, 24, 0, 0, 0, and 1 death of male mice, and 1, 0, 24, 1, 0, 1, and 2 death of female mice in vehicle control 1, vehicle control 2, positive control, low, medium, mid-hi, and high dose groups, respectively. Tests did not show statistically significant dose response relationship in mortality across combined control, low, medium, mid-hi, and high dose groups in either sex. The pairwise comparisons show statistically significant increased mortality in male mice control 2 compared to control 1. The pairwise comparisons also showed statistically significant increased mortality in the positive control group

compared to both control 1 and control 2 in both sexes.

**Reviewer’s comment:** *The sponsor’s calculation showed one death in the female medium dose group, and this reviewer’s calculation showed no deaths in this group. This difference is due to the fact that one female mouse (# 100261) from medium dose group died due to natural causes during the terminal sacrifice weeks. The sponsor counted this with the natural deaths, while this reviewer counted this with the terminally sacrificed mice.*

**3.2.2. Tumor data analysis**

The tumor rates and the p-values of the tested tumors are listed in Tables 6A and 6B in the appendix for male and female mice, respectively.

**Reviewer’s findings:** The following tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of combined control and treated groups.

**Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups With Combined Control in Mice**

Female mice							P-Value				
Organ Name	Tumor Name	Com C <sup>#</sup> N=50	Low N=25	Med N=25	Mid-Hi N=25	High N=25	Dose Resp	Com C vs L	Com C vs M	Com C vs MH	Com C vs H
Whole body	Hemangi osarcoma	0	1	2	1	4	0.0085*	0.3378	0.1111	0.3288	0.0110*

<sup>#</sup> Com C: Combined vehicle control

Based on the criteria of adjustment for multiple testing discussed in the rat data analysis section, the incidence of whole body hemangiosarcoma was considered to have a statistically significant dose response relationship in female mice. The pairwise comparisons showed statistically significant increase incidence of whole body hemangiosarcoma in female mice high dose group compared to the combined control.

Tests showed statistically significant increased incidences of lung adenoma, adenocarcinoma, and hemangiosarcoma, spleen hemangiosarcoma, and whole body hemangiosarcoma in the positive control compared to the combined control in both sexes. Results of this analysis are given in Table 7A and 7B in the appendix for male and female mice.

**4. Summary**

In this submission the sponsor included reports of two animal carcinogenicity studies, one in regular rats and one in transgenic mice. These studies were intended to assess the carcinogenic potential of MK4305 (Suvorexant) when administered orally by gavage once daily at appropriate drug levels for 104 weeks in rats and for 26 weeks in mice.

In this review the phrase "dose response relationship" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

**Rat Study:** Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were three treated groups and two identical vehicle control groups. Two hundred and fifty Sprague -Dawley, CrI:CD(SD) rats of each sex were randomly allocated to treated and control groups in equal

size of 50 rats. The dose levels for treated groups were 80, 160, or 325 mg/kg/day for males and 40, 80, or 325 mg/kg/day for females. Control 1 and Control 2 rats were given 1300 mg/kg/day of HPMC (hydroxypropyl methylcellulose-acetate succinate) in 0.5% methylcellulose with 5 mM hydrochloric acid in deionized water (vehicle) by gavage.

During the administration period all rats were observed daily for mortality and morbidity. All rats were observed weekly for clinical signs through Study Week 104. Beginning in Study Week 26, all rats were palpated for masses every 4 weeks through Study Week 104. Additionally, unscheduled palpations were performed on rats found with masses in Study Weeks 6, 10, 14, 18, 23, 39, 44, 52, 53, 59, 60, 61, 64, 68, and generally weekly thereafter. Body weights for all rats were measured at pretest, once per week from Study Week 1 through Study Week 14, and once every 4 weeks thereafter through Study Week 102.

The tests did not show statistically significant dose response relationship in mortality across combined control and treated groups in either sex. The pairwise comparisons showed statistically significant decreased mortality in male medium dose group compared to combined control.

The tests showed statistically significant positive dose response relationship in the incidence of hepatocellular adenoma, and combined incidences of hepatocellular adenoma and carcinoma in male rats, and incidence of thyroid follicular cell adenoma, combined incidences of thyroid follicular cell adenoma and carcinoma in both sexes. The pairwise comparisons showed statistically significant increased incidence of hepatocellular adenoma, combined incidences of hepatocellular adenoma and carcinoma in the male rat high dose group compared to the combined control. The pairwise comparisons also showed statistically significant increased incidence of thyroid follicular cell adenoma, and combined incidences of thyroid follicular cell adenoma and carcinoma in male rat medium and high dose groups, and in female rat high dose group compared to the combined control.

**Mouse Study:** Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were four treated groups, two vehicle control groups, and one positive control group. One hundred and seventy five male and female ic:CB6F1-TgrasH2@Jcl mice were assigned randomly to the treated, vehicle control and positive control groups in equal size of 25 mice per group. The dose levels for treated groups were 25, 50, 200, and 650 mg/kg/day. The vehicle controls received the vehicle (sterile water for injection) by gavage. Positive control mice were dosed via intraperitoneal (i.p.) injections of urethane (L-000471271) in saline on Days 1, 3 and 5, at a dose level of 1000 mg/kg/day. A dose volume of 10 mL/kg body weight was used for all groups.

All mice except those from the positive control group were observed daily for mortality, and once a week for physical signs through Study Week 26. All mice from the positive control group were observed daily for mortality, and on the dosing days, once per week from Study Week 2 to Study Week 26 for physical signs. All mice except those from the positive control group were palpated for masses every 4 weeks, from Study Week 14 to Study Week 26. Body Weights of all mice were measured pretest and once per week from Study Week 1 through Study Week 26.

Tests did not show statistically significant dose response relationship in mortality across combined control, low, medium, mid-hi, and high dose groups in either sex. The pairwise comparisons show statistically significant increased mortality in male control 2 compared to control 1. The pairwise comparisons also showed statistically significant increased mortality in the positive control group compared to both control 1 and control 2 in both sexes.

Tests showed statistically significant dose response relationship in the incidence of whole body hemangiosarcoma in female mice. The pairwise comparisons showed statistically significant increase incidence of whole body hemangiosarcoma in female mice high dose group compared to the combined control.

Tests showed statistically significant increased incidences of lung adenoma, adenocarcinoma, and hemangiosarcoma, spleen hemangiosarcoma, and whole body hemangiosarcoma in the positive control compared to the combined control in both sexes.

Mohammad Atiar Rahman, Ph.D.  
Mathematical Statistician

Concur: Karl Lin, Ph.D.  
Team Leader, Biometrics-6

cc:  
Archival NDA 204-569  
Dr. Siarey  
Ms. Michaloski

Dr. Machado  
Dr. Lin  
Dr. Rahman  
MS. Patrician

5. Appendix

**Table 1A: Intercurrent Mortality Rate  
Male Rats**

Week	0 mg kg day		0 mg kg day		80 mg kg day		160 mg kg day		325 mg kg day	
	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	4	8.00	4	8.00	3	6.00	.	.	3	6.00
53 - 78	5	18.00	8	24.00	5	16.00	3	6.00	3	12.00
79 - 91	5	28.00	5	34.00	7	30.00	.	.	6	24.00
92 - 104	6	40.00	9	52.00	9	48.00	11	28.00	9	42.00
Ter. Sac.	30	60.00	24	48.00	26	52.00	36	72.00	29	58.00
Total	N=50		N=50		N=50		N=50		N=50	

# Cum. %: Cumulative percentage

**Table 1B: Intercurrent Mortality Rate  
Female Rats**

Week	0 mg kg day		0 mg kg day		40 mg kg day		80 mg kg day		325 mg kg day	
	No. of Death	Cum. %	No. of Death	Cum. %						
0 - 52	.	.	3	6.00	.	.	1	2.00	1	2.00
53 - 78	4	8.00	8	22.00	4	8.00	2	6.00	2	6.00
79 - 91	3	14.00	6	34.00	3	14.00	6	18.00	2	10.00
92 - 104	9	32.00	6	46.00	10	34.00	5	28.00	9	28.00
Ter. Sac.	34	68.00	27	54.00	33	66.00	36	72.00	36	72.00
Total	N=50		N=50		N=50		N=50		N=50	

# Cum. %: Cumulative percentage

**Table 2A: Intercurrent Mortality Comparison  
Male Rats**

Test	Statistic	P_Value
Dose-Response	Likelihood Ratio	0.2586
Homogeneity	Log-Rank	0.1002

#P-Values were calculated using data from Combined vehicle control, Low, Medium, and High dose groups

**Table 2B: Intercurrent Mortality Comparison  
Female Rats**

Test	Statistic	P_Value
Dose-Response	Likelihood Ratio	0.4929
Homogeneity	Log-Rank	0.4200

#P-Values were calculated using data from Combined vehicle control, Low, Medium, and High dose groups









**Table 4A: Intercurrent Mortality Rate in Male Mice**

Week	Control 1		Control 2		Positive Control		Low		Medium		Mid-Hi		High	
	No. of		No. of		No. of		No. of		No. of		No. of		No. of	
	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %
0 - 10	.	.	1	4.00	1	4.00	.	.	.	.	.	.	.	.
11 - 20	.	.	.	.	15	64.00	.	.	.	.	.	.	.	1
21 - 26	.	.	3	16.00	8	96.00	.	.	.	.	.	.	.	.
Ter. Sac.	25	100.00	21	84.00	1	4.00	25	100.00	25	100.00	25	100.00	24	96.00
Total	N=25		N=25		N=25		N=25		N=25		N=25		N=25	

# Cum. %: Cumulative percentage

**Table 4B: Intercurrent Mortality Rate Female Mice**

Week	Control 1		Control 2		Positive Control		Low		Medium		Mid-Hi		High	
	No. of		No. of		No. of		No. of		No. of		No. of		No. of	
	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %
11 - 20	.	.	.	.	11	44.00	.	.	.	.	1	4.00	2	8.00
21 - 26	1	4.00	.	.	13	96.00	1	4.00	.	.	.	.	.	.
Ter. Sac.	24	96.00	25	100.00	1	4.00	24	96.00	25	100.00	24	96.00	23	92.00---
Total	N=25		N=25		N=25		N=25		N=25		N=25		N=25	

# Cum. %: Cumulative percentage

**Table 5A: Intercurrent Mortality Comparison Male Mice**

Test	Statistic	P_Value#
Dose-Response	Likelihood Ratio	0.8142
Homogeneity	Log-Rank	0.2022

#P-Values were calculated using data from Combined vehicle control, Low, Medium, Mid-Hi, and High dose groups

**Table 5B: Intercurrent Mortality Comparison Female Mice**

Test	Statistic	P_Value#
Dose-Response	Likelihood Ratio	0.2855
Homogeneity	Log-Rank	0.8212

#P-Values were calculated using data from Combined vehicle control, Low, Medium, Mid-Hi, and High dose groups

**Table 6A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons  
Male Mice**

Organ Name	Tumor Name	0 mg Com C N=50	25 mg Low N=25	50 mg Med N=25	200 mg Mid-Hi N=25	650 mg High N=25	P_Val ue Dose Resp	P_Val ue Com C vs L	P_Val ue Com C vs M	P_Val ue Com C vs MH	P_Val ue Com C vs H
Body Cavi ties/P	Hemangi osarcoma	1	0	0	0	0	0. 6735	0. 3425	0. 3425	0. 3425	0. 3333
	Mesothel ioma	1	0	0	0	0	0. 6735	0. 3425	0. 3425	0. 3425	0. 3333
Harder ian Gl and	Adenoma	1	1	1	1	1	0. 3342	0. 5708	0. 5708	0. 5708	0. 5587
Li ver	Hepatocel lular adenoma	0	0	1	0	0	0. 3333	.	0. 3425	.	.
Lung	Adenocarci noma	1	0	0	0	0	0. 6735	0. 3425	0. 3425	0. 3425	0. 3333
	Adenoma	0	0	1	0	1	0. 1375	.	0. 3425	.	0. 3333
Skel etal Muscl e	Hemangi osarcoma	0	0	0	1	0	0. 3333	.	.	0. 3425	.
Ski n	Fi brosarcoma	1	0	0	0	0	0. 6735	0. 3425	0. 3425	0. 3425	0. 3333
	Hemangi osarcoma	1	0	0	0	0	0. 6735	0. 3425	0. 3425	0. 3425	0. 3333
Spl een	Hemangi osarcoma	2	1	1	1	0	0. 7692	0. 2685	0. 2685	0. 2685	0. 5587
Stomach/Nongl an	Papi l loma	0	1	0	0	1	0. 1935	0. 3425	.	.	0. 3333
Whol e body	Hemagi osarcoma	4	1	1	2	0	0. 8658	0. 5621	0. 5621	0. 3335	0. 8109

-----  
Com C: Combined vehi cle control

**Table 6B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons Female Mice**

Organ Name	Tumor Name	0 mg Com C N=50	25 mg Low N=25	50 mg Med N=25	200 mg Mid-High N=25	650 mg High N=25	P_Val ue Dose Resp	P_Val ue Com C vs L	P_Val ue Com C vs M	P_Val ue Com C vs MH	P_Val ue Com C vs H
Body Cavi ties/B	Hemangi osarcoma	0	0	0	0	1	0.1633	.	.	.	0.3288
Harderian GI and	Adenoma	1	2	0	0	0	0.8191	0.2622	0.3378	0.3288	0.3194
Kidney	Adenoma	0	0	0	1	0	0.3219	.	.	0.3288	.
	Hemangi osarcoma	0	1	0	0	0	0.4932	0.3378	.	.	.
Lung	Adenocarci noma	0	1	0	0	0	0.4932	0.3378	.	.	.
	Adenoma	0	0	1	0	0	0.3219	.	0.3378	.	.
Primary Site Un	Lymphoma	0	0	1	0	0	0.3219	.	0.3378	.	.
Skin	Hemangi osarcoma	0	0	0	0	1	0.1633	.	.	.	0.3288
Spleen	Hemangi osarcoma	0	0	2	1	1	0.1633	.	0.1111	0.3288	0.3194
Stomach/Nongl an	Papi l l oma	1	0	0	0	0	0.6644	0.3378	0.3378	0.3288	0.3194
	Squamous Cell Carci noma	0	0	0	1	0	0.3219	.	.	0.3288	.
Uterus	Pol yp	0	1	0	0	0	0.4932	0.3378	.	.	.
Vagina	Hemangi osarcoma	0	0	0	0	1	0.1575	.	.	.	0.3194
Whole_Body	Hemagi osarcoma	0	1	2	1	4	0.0085*	0.3378	0.1111	0.3288	0.0110*

-----  
Com C: Combined vehicle control

**Table 7A: Pairwise Comparisons of Positive Control and Vehicle Groups  
Male Mice**

Organ Name	Tumor Name	0 mg	25 mg	P_Val ue
		Com C# N=50	Pos C# N=25	Com C vs Pos C
fff				
Body Cavi ties/P	Hemangi osarcoma	1	0	0.1724
	Mesothel ioma	1	0	0.1724
Harderian Gl and	Adenoma	1	0	0.1724
Lung	Adenocarci noma	1	14	<0.001*
	Adenoma	0	11	<0.001*
	Hemangi osarcoma	0	2	0.0272*
Ski n	Fi brosarcoma	1	0	0.1724
	Hemangi osarcoma	1	0	0.1724
Spl een	Hemangi osarcoma	2	21	<0.001*
Whol e_Body	Hemagi osarcoma	4	22	<0.001*

-----  
 # Com C: Combi ned vehi cle control ; Pos C: Posi ti ve control



Figure 1A: Kaplan-Meier Survival Functions for Male Rats

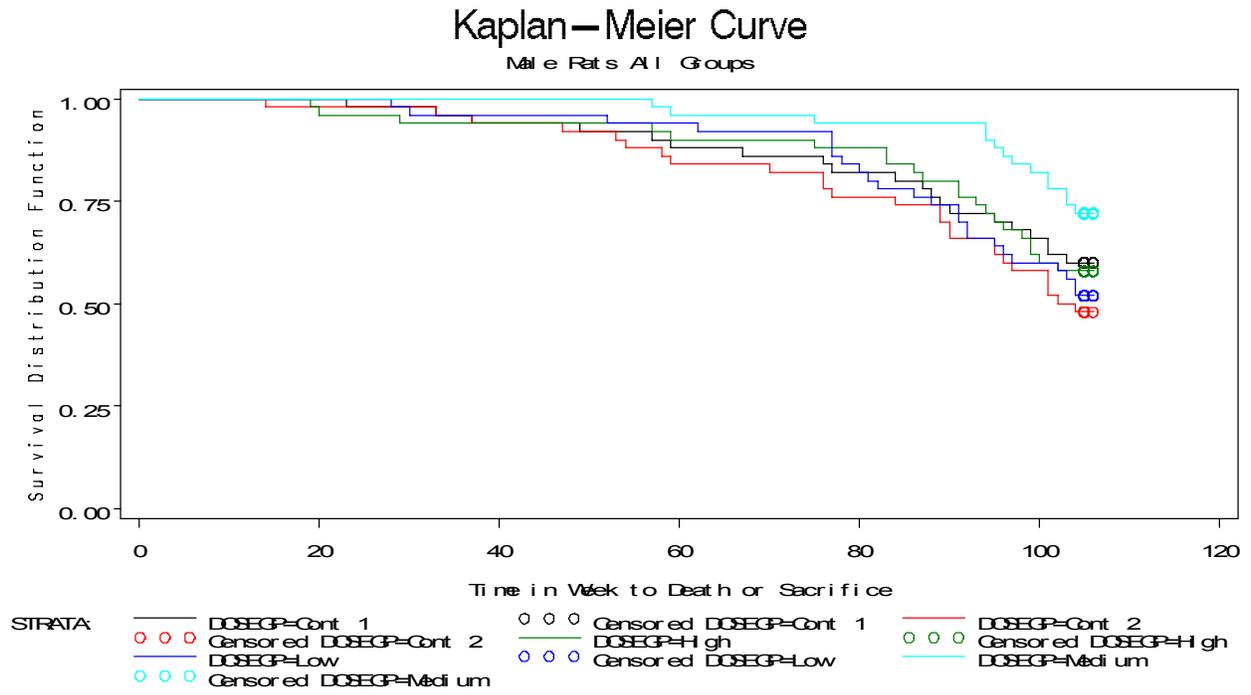


Figure 1B: Kaplan-Meier Survival Functions for Female Rats

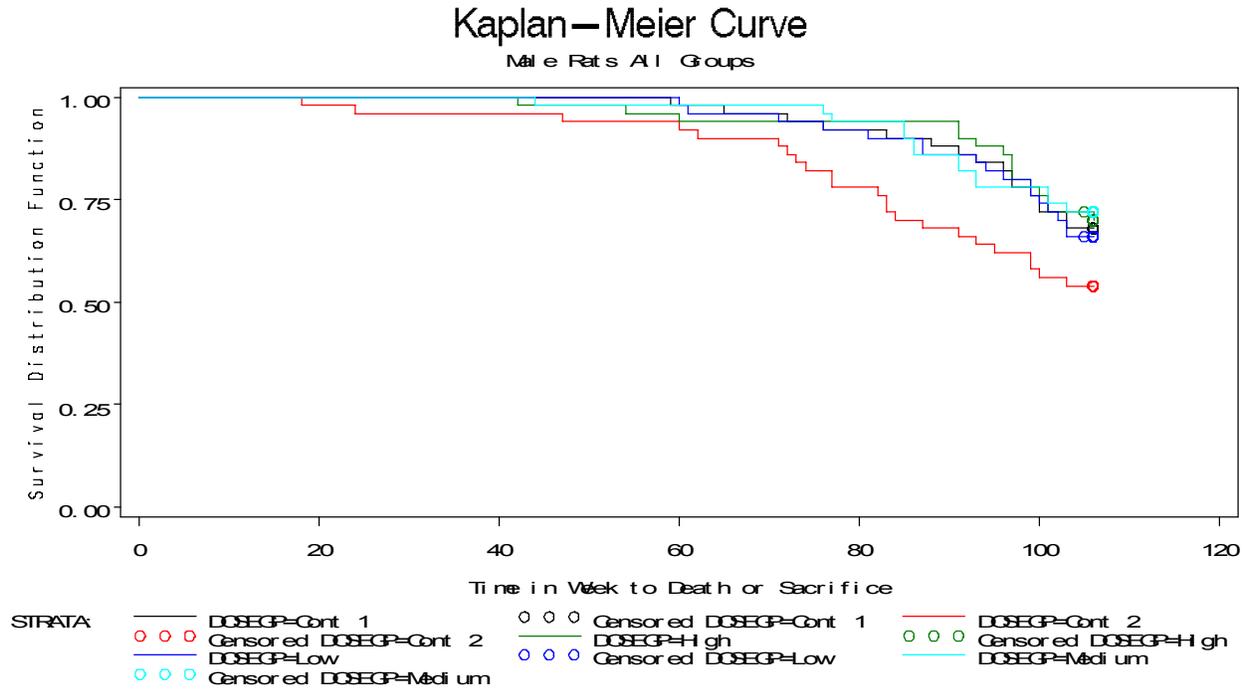


Figure 2A: Kaplan-Meier Survival Functions for Male Mice

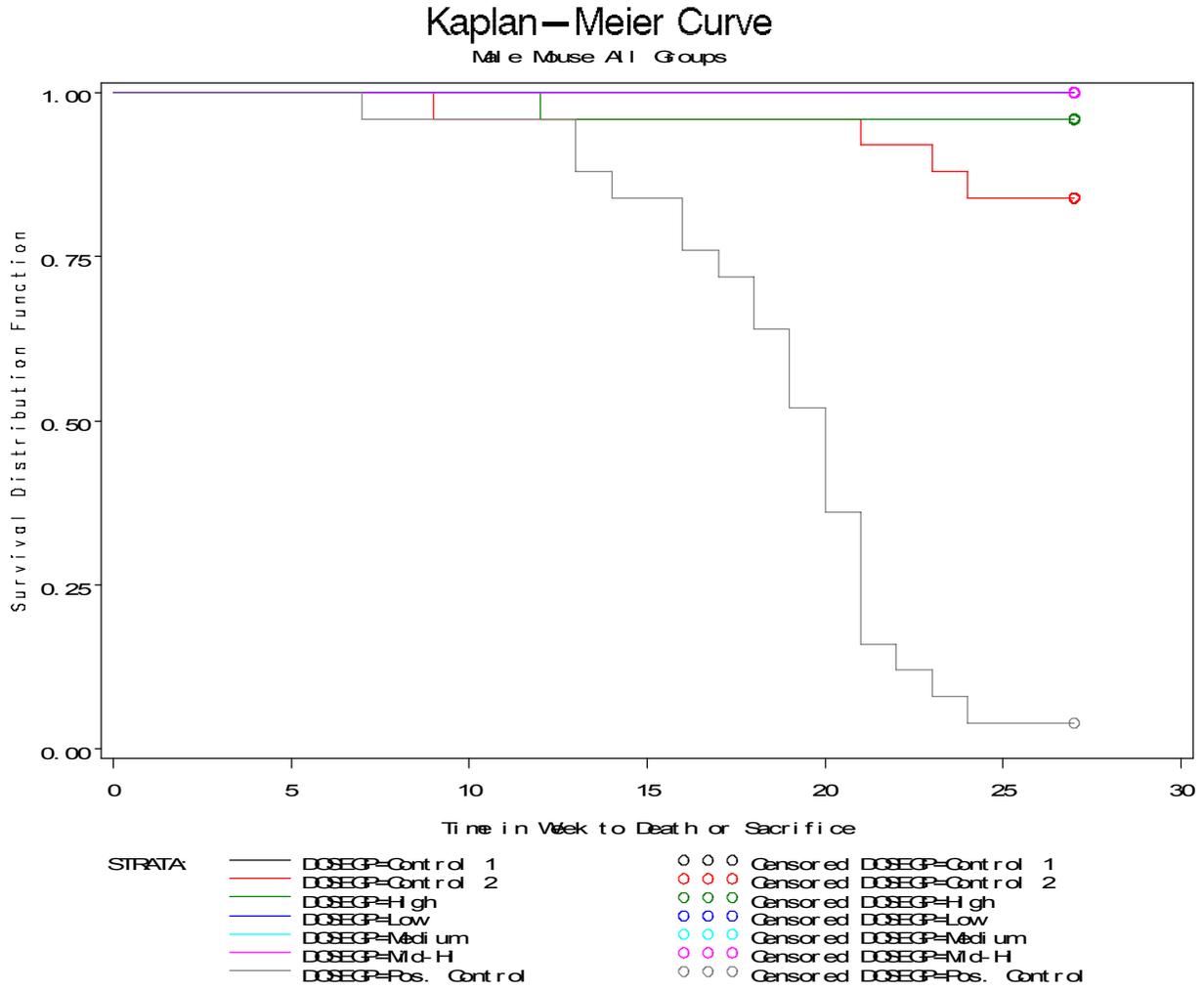
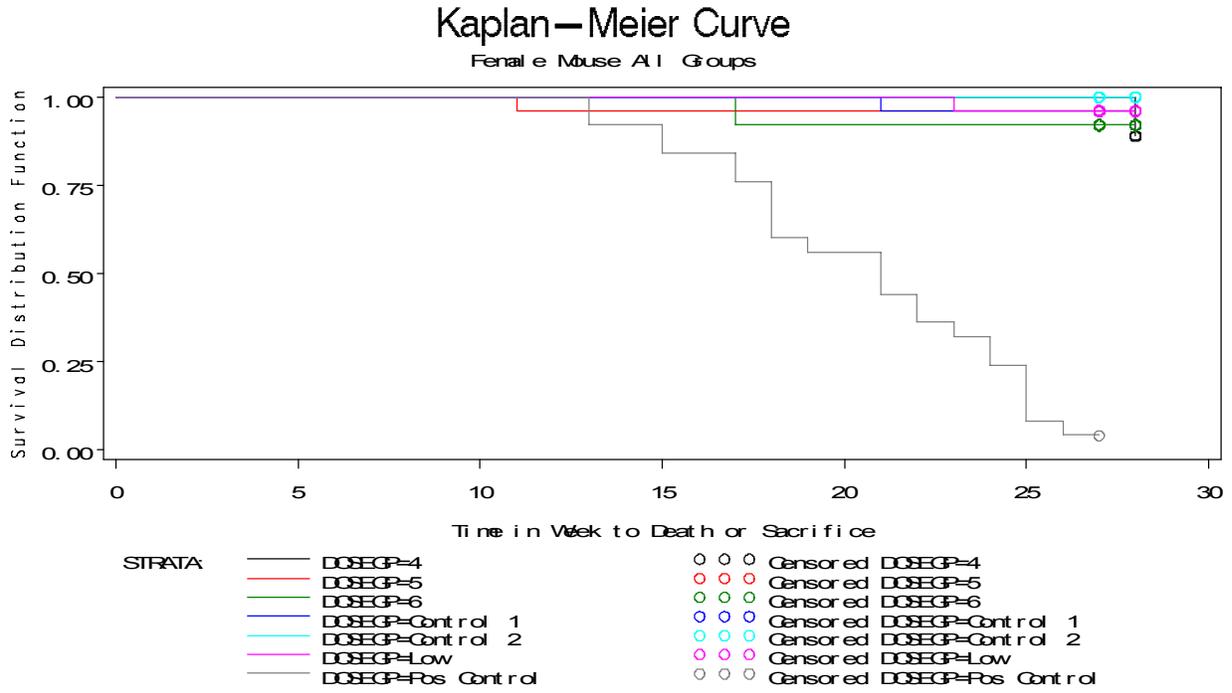


Figure 2B: Kaplan-Meier Survival Functions for Female Mice



## References

1. Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J. Wahrendorf, "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva*, 311-426, 1980.
2. Bailer AJ, Portier CJ (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples." *Biometrics*, 44, 417-431.
3. Bieler, G. S. and Williams, R. L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.
4. Tarone RE, "Test for trend in life table analysis", *Biometrika* 1975, 62: 679-82
5. Lin K.K. and Rahman M.A., "Overall false positive rates in tests for linear trend in tumor incidence in animal carcinogenicity studies of new drugs", *Journal of Biopharmaceutical Statistics*, 8(1), 1-15, 1998.
6. Haseman, J, "A re-examination of false-positive rates for carcinogenesis studies", *Fundamental and Applied Toxicology*, 3: 334-339, 1983.
7. Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), May 2001.

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

MOHAMMAD A RAHMAN  
01/10/2013

KARL K LIN  
01/10/2013  
Concur with review