

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*  
**022328Orig1s000**

**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:** 23228 (S0036)

**Drug Name:** Zolpidem Sublingual

**Indication(s):** Insomnia: Middle of the Night Awakening

**Applicant:** Transcept

**Date(s):** January 14, 2011

**Review Priority:** Standard (Resubmission)

**Biometrics Division:** Division of Biometrics 1

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

**Concurring Reviewers:** Kun Jin, Ph.D., Team Leader

Hsien Ming (James) Hung, Ph.D., Director, Division of Biometrics 1

**Medical Division:** Division of Neurology

**Clinical Team:** Carole Davis, M.D.

Ron Farkas, M.D., Ph.D., Team Leader

**Project Manager:** Cathleen Michaloski

**Keywords:**

Crossover Design, Symmetry Analysis, Prespecified Data Editing Procedure, Safety, Multiplicity

# Table of Contents

STATISTICAL REVIEW AND EVALUATION .....	1
LIST OF TABLES.....	3
LIST OF FIGURES.....	4
1. EXECUTIVE SUMMARY .....	5
2. INTRODUCTION .....	6
2.1 OVERVIEW.....	6
2.2 DATA SOURCES .....	8
3. STATISTICAL EVALUATION .....	8
3.1 DATA AND ANALYSIS QUALITY .....	8
3.2 EVALUATION OF EFFICACY .....	8
3.3 EVALUATION OF SAFETY .....	9
3.3.1 <i>Study ZI-18</i> .....	9
3.3.1.1 Study Design and Analysis Plan .....	9
3.3.1.2 Baseline Demographics.....	14
3.3.1.3 Randomly Assigned Crossover Sequences of Treatment .....	15
3.3.1.4 Subject Disposition .....	18
3.3.1.5 Results.....	18
3.3.1.6 Reviewer’s Analysis of All Driving Data Including “Artifacts”.....	27
3.3.1.7 Additional Reviewer Comments on the SDLP as a measure of impaired driving.....	28
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....	29
4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION .....	29
4.1.1 <i>Gender</i> .....	29
4.1.2 <i>Race</i> .....	30
4.1.3 <i>Age</i> .....	30
4.1.4 <i>US vs. non-US</i> .....	32
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS .....	32
5. SUMMARY AND CONCLUSIONS .....	32
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	32
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	33

## LIST OF TABLES

Table 1 Timing of Treatment and Driving Test.....	10
Table 2 Baseline Demographics .....	15
Table 3 Observed Frequency of Randomized Sequences of Treatment .....	16
Table 4 Observed Frequency of Actual Sequences of Treatment.....	17
Table 5 Summary Statistics of SDLP for the Two Placebo Treatments .....	18
Table 6 Number Impaired SDLP Changes of Active Control from Placebo .....	19
Table 7 Number of ZST 4hr Changes from Placebo exceeding Various SDLP Cutpoints.....	20
Table 8 Number of ZST 3hr Changes from Placebo exceeding Various SDLP Cutpoints.....	23
Table 9 Secondary Analysis of SDLP as a Continuous Variable (No cutpoint).....	24
Table 10 Differences in SDLP between ZST 3hr and ZST 4 hr .....	27
Table 11 Differences from placebo in SDLP with all of raw data, i.e., no artifacts removed .....	27
Table 12 Differences between ZST 3hour and Placebo Exceeding Various Thresholds by Gender .....	29
Table 13 Differences between ZST 4hour and Placebo Exceeding Various Thresholds by Gender .....	30
Table 14 Differences between ZST 3hour and Placebo Exceeding Various Thresholds by Age Group .....	31
Table 15 Differences between ZST 4hour and Placebo Exceeding Various Thresholds by Age Group .....	31
Table 16 Differences between ZOP (Active Control) and Placebo Exceeding Various Thresholds by Age Group.....	32

## LIST OF FIGURES

Figure 1 Plot of SDLPs During ZST 4hr Treatment versus Placebo Treatment.....	21
Figure 2 Cumulative Distribution of SDLP.....	22
Figure 3Cumulative Distribution of Change from Placebo in SDLP .....	25
Figure 4 SDLP Difference between ZST 4 hr and Placebo .....	26

## 1. EXECUTIVE SUMMARY

Efficacy of treating middle of the night awakening was evaluated in a previous submission. This submission focused on next day driving safety. A single crossover designed study of next day driving safety after treatment with Zolpidem sublingual (ZST) demonstrated assay sensitivity by means of the comparison of the active control, Zopiclone, to placebo. The primary endpoint was the standard deviation of lateral position (SDLP) on the road as obtained from a standardized driving test. The primary analysis involved the number of driving impaired subjects, defined as those having an increase from placebo in SDLP under drug treatment that exceeded 2.5 cm. There is some question about the most relevant cutoff of the distribution of differences from placebo in SDLP used to define impairment so the sponsor also prespecified several other cutoffs. Because this is a safety study there was no multiplicity adjustment applied for examining more than one cutoff. In this analysis the number of impaired subjects was compared to the number of improved subjects, defined as those having a decrease from placebo in SDLP under drug treatment that was below -2.5 cm. The SDLP data suggests that ZST 3 hour, i.e., taking ZST 3 hours before awakening, impaired next day driving. Although ZST 4 hour did not reach nominal significance for the primary cutoff of 2.5 cm, or several other cutoffs specified by the sponsor, there were still other cutoffs in addition to the prespecified secondary analysis of mean differences which suggested impairment for ZST 4 hour compared to placebo. There is a multiplicity issue related to using several cutoffs but we are dealing with a potentially very serious safety issue and a questionable cutoff designated as primary. Thus, overall, it seems to this reviewer that there may be some impairment of next day driving caused by taking ZST 4 hours before rising as well.

This study utilized a four period crossover design. Two dosing times, 3 hours and 4 hours before rising time, were investigated but the sponsor decided that it was not practical or necessary to have each subject undergo the six periods of scientific interest (investigational 3 and 4 hour, placebo 3 and 4 hour, and active control 3 hour and 4 hour). Therefore, all subjects underwent both 3 hour and 4 hour for ZST in different periods but they underwent only one time for placebo and only one time for active control. If it turned out that placebo 3 hours before rising and placebo 4 hours before rising had different effects on next day driving then the study may have been underpowered.

There was a prespecified procedure for having trained, blinded personnel edit the continuously sampled driving data prior to analysis in order to remove anticipated data “artifacts” of the driving test, such as passing slow cars, not enough light entering the camera, etc. Although excluding data always arouses suspicions it was prespecified in this case for legitimate reasons associated with conducting the driving test on the open road and, ostensibly, proper blinding of the editors was used. Of course, there is some risk of misclassification. However, it is important to distinguish between truly impaired driving and permitted driving behavior that could appear as impaired, such as passing a slow car. Therefore, it seems justified in this case and we will rely on the “artifact” free driving data. This would probably have been more of a concern if the investigational drug had shown no evidence of impairment in the edited data but showed

evidence in the unedited data. Both forms of the data were examined and this was found not to be the case.

The sponsor's primary analysis involving a statistical test of symmetry, i.e., comparing the proportion impaired to the proportion "improved" may not always make clinical sense. In particular, if there is a well established SDLP cutoff corresponding to impairment and  $\pi_{I+} = \pi_{I-}$ , i.e., the proportion of impaired drivers equals the proportion of "improved" drivers, and  $\pi_{I+}$  is moderately large then even though the distribution is symmetric there is a moderately large proportion of impaired drivers. Ignoring this would be like ignoring a common serious adverse event which we would not do. Therefore, without further constraints on the size of the proportion impaired this symmetry test seems problematic from a regulatory perspective in this setting of driving safety. In other words, the sponsor's test may mathematically cancel out a serious clinical issue between two different groups, but the impaired group still exists. This is a cause for concern about the use of this symmetry test in this type of study.

Single site studies such as this may reduce variability compared to multi-site studies but may have less generalizability.

## 2. INTRODUCTION

Efficacy of this drug was concluded in a review of a prior NDA submission for this drug. Because of a clinical concern about next day effects due to the novelty of the proposal for indicating this drug to be used in the middle of the night a driving safety study was required. This submission involves the driving study, ZI-18.

### 2.1 Overview

This driving safety study, ZI-18, was a single center foreign double-blind crossover designed study done at Maastricht University in the Netherlands. Forty healthy volunteers were randomized. Each received four treatments sequentially, according to randomized sequence, and each took a driving test the day after each treatment.

As more than one subject was evaluated each day, times were staggered by 5-10 minutes so that each subject could be treated individually. Treatment periods for a given subject were separated by at least 3 days.

On March 24, 2010 there was a teleconference with the sponsor on the plan for the driving study. Biometrics expressed some reservations about the test of symmetry. For example, since it suggests under the null hypothesis that the drug improves driving for some, in fact, as many as it impairs. Also, because the focus of this study's focus is safety the issue of non-inferiority being a

more appropriate testing approach was raised. However, the sponsor said that the required sample size for a non-inferiority study would not be practical. Also, there is no established non-inferiority margin. Therefore, the Division did not reject the test of symmetry being the primary analysis method.

The sponsor also submitted a meta-analysis of driving safety studies (ECA-001). This was not reviewed in detail because the meta-analysis dataset was not included with the initial submission of the NDA supplement. It was received 2 months later after a request. It doesn't change the driving safety issue observed for ZST 3 hr in study ZI-18. Furthermore, the mean placebo SDLP in study ZI-18 was numerically smaller (from 1 up to 6 cm) than that in any of the other 8 studies involved in the meta-analysis which raises questions about comparability of ZI-18 and the other studies.



## 2.2 Data Sources

The following dataset contains the SDLP data summary of the individual patients' driving data for the primary analysis.

<\\Cdsub1\evsprod\NDA022328\0033\m5\datasets\zi-18\analysis\adxd.xpt>

The sponsor submitted this data in SDTM format along with analysis programs.

The sponsor's study report can be found here.

<\\Cdsub1\evsprod\NDA022328\0033\m5\53-clin-stud-rep\534-rep-human-pd-stud\5341-healthy-subj-pd-stud-rep\zi-18\zi-18-body.pdf>

Upon request after the initial submission the sponsor submitted the raw data in separate files for each of the 40 patients.

[\\Cdsub1\evsprod\NDA022328\0037\m5\datasets\zi-18\tabulations \(b\) \(4\) \drvst01.xpt](\\Cdsub1\evsprod\NDA022328\0037\m5\datasets\zi-18\tabulations\ (b) (4) \drvst01.xpt)

...

[\\Cdsub1\evsprod\NDA022328\0037\m5\datasets\zi-18\tabulations \(b\) \(4\) \drvst40.xpt](\\Cdsub1\evsprod\NDA022328\0037\m5\datasets\zi-18\tabulations\ (b) (4) \drvst40.xpt)

The datasets for the driving safety meta-analysis described in report ECA-001 are at the following locations.

<\\Cdsub1\evsprod\NDA022328\0036\m5\datasets\eca-001\analysis\admeans.xpt>

<\\Cdsub1\evsprod\NDA022328\0036\m5\datasets\eca-001\analysis\adsdlpd.xpt>

The meta-analysis report is located as follows.

<\\Cdsub1\evsprod\NDA022328\0033\m5\54-lit-ref\laska-2010-report-eca-001.pdf>

## 3. STATISTICAL EVALUATION

### 3.1 Data and Analysis Quality

### 3.2 Evaluation of Efficacy

The study under consideration, ZI-18, is focused on driving safety on the morning after treatment. Efficacy was examined in the review of the earlier NDA submission of 2008.

### **3.3 Evaluation of Safety**

#### **3.3.1 Study ZI-18**

Only safety related to next day driving, as determined solely from the structured driving study ZI-18, is evaluated in this section. For a more general safety assessment please see the clinical review.

##### **3.3.1.1 Study Design and Analysis Plan**

###### **Primary objective**

The primary objective of this study was to assess the risk of impaired driving in the morning at 3 and 4 hours after a middle of the night dose of zolpidem tartrate sublingual tablet 3.5 mg.

###### **Secondary objectives**

The secondary objective of this study was to assess aspects of driving performance as they relate to absolute standard deviation of lateral position (SDLP) and standard deviation of speed (SDS).

This was a single-center, double-blind, randomized, placebo-controlled, four-way crossover study. Upon giving informed consent, 40 adult male and female healthy subjects were enrolled in the study and randomized to one of 24 possible treatment sequences of 4 treatment periods.

The four treatment conditions in the study were:

- A. Zopiclone oral capsule 7.5 mg administered 9 hours before driving, as positive control (ZOP)
- B. Zolpidem sublingual tablet 3.5 mg administered 3 hours before driving (ZST 3h)
- C. Zolpidem sublingual tablet 3.5 mg administered 4 hours before driving (ZST 4h)
- D. Placebo

The time schedule for each treatment is shown in Table 1.

**Table 1 Timing of Treatment and Driving Test**

<b>Treatment condition:</b>	<b>Bedtime dose</b>		<b>MOTN dose</b>		<b>Rise time</b>	<b>Test</b>
<b>A1</b>	2315h	ZOP	0515h	Placebo	0730h	0815h
<b>A2</b>	2315h	ZOP	0415h	Placebo	0730h	0815h
<b>B 3h</b>	2315h	Placebo	0515h	ZST	0730h	0815h
<b>C 4h</b>	2315h	Placebo	0415h	ZST	0730h	0815h
<b>D1</b>	2315h	Placebo	0515h	Placebo	0730h	0815h
<b>D2</b>	2315h	Placebo	0415h	Placebo	0730h	0815h
*As more than one subject will be evaluated per visit, times will be staggered by approximately 10 minutes so that each subject can be treated individually						

Note: Table copied from page 26 of protocol

The randomization procedure was based on Williams design. A Williams design is a (generalized) Latin square that is also balanced for first order carryover effects in that each treatment appears in every row and every column exactly once and every treatment is preceded by every other treatment exactly once (Sharma, 1975; Williams, 1949).

Study medication was administered at bedtime, which was approximately at 23:15 hours. At approximately 04:15 or 05:15 hours, subjects were awakened and a second study medication was administered. Approximately half the subjects in Treatment Condition A (N = 40) were in A1 (N=19) and the remainder in A2 (N = 21). Similarly, approximately half of the subjects in D (N = 40) were in D1 (N = 21) and the remainder in D2 (N = 19).

### Primary endpoint

The primary endpoint was to be Impaired Drivers, where Impaired Driver (ImpDr) is a binary indicator defined for subject  $i$  to be 1 if  $\text{TreatmentSDLP}[i] - \text{PlaceboSDLP}[i] > 2.5$  cm and zero otherwise.

### Secondary endpoints

The secondary endpoints were to be Impaired Drivers, where Impaired Driver (ImpDr) is a binary indicator defined for subject  $i$  to be 1 if:

- $\text{TreatmentSDLP}[i] - \text{PlaceboSDLP}[i] > 2$  cm and zero otherwise,
- $\text{TreatmentSDLP}[i] - \text{PlaceboSDLP}[i] > 3.5$  cm and zero otherwise, and

Additional secondary endpoints were to be SDLP in centimeters in the highway driving test and standard deviation in speed (SDS).

### Methods of Assessment for Outcome Variables

SDLP and SDS are objective measurements obtained during the highway driving test using validated instruments attached to the automobile. Subjects were to operate a vehicle equipped with validated instruments which continuously record speed and lateral position. Subjects were to be instructed to drive in the center of the slower traffic lane while maintaining a constant speed of 95km/hour over a 100-km area of highway. SDLP is the primary endpoint for the standardized Highway Driving Test (O'Hanlon, 1984) and describes the variation (standard deviation) from

the mean lateral position on the road.

## STATISTICS

The Intent-to-treat (ITT) analysis set was defined as all randomized subjects who received at least one dose of study drug. All ITT subjects were to be summarized, based on their randomized treatment group, for primary and secondary analyses, subject disposition, demographics, baseline characteristics, as well as included in listings.

The safety analysis set was defined as all subjects who received at least one dose of study drug. The subjects were to be summarized based on their actual treatment received in the Safety analysis set. All safety analyses were to be based on this set.

The ITT analysis sets were to be used in both the primary and secondary analyses.

The primary outcome measure was a binary variable indicating that a difference between treatment and placebo in SDLPs exceeds the threshold of 2.5. As a sensitivity test of the dependence of the results on the choice of threshold value, several cutpoints were to also be utilized. In particular, the sponsor proposed to use 2.0 and 3.5 as two additional fixed cutpoints. Secondary measures that are binary were to be analyzed in the same fashion.

The remaining outcome measures from the highway driving test were the SDLP in cms and the standard deviation of speed (SDS).

## Formulating the null hypothesis for Impaired Driving

Suppose that a threshold,  $t^*$ , has been determined and that the proportion of the  $N$  subjects in the study whose change on ZST from the placebo condition exceeds  $t^*$  is calculated. This calculation assumes that the two placebo groups D1 and D2 have been combined into one group.

With  $I$  indicating ZST, let

$n_I^-$  = number of subjects such that  $Y_I - Y_p > t^*$

$p_I^- = n_I^- / N$  = observed proportion of subjects whose driving worsened on ZST

Then  $p_I^-$  is an estimator of the population probability  $\pi_I^-$  that driving is impaired after using ZST.

Let

$n_I^+$  = number of subjects such that  $Y_I - Y_p < -t^*$

$p_I^+ = n_I^+ / N$  = observed proportion of subjects whose driving improved on ZST

Then  $p_I^+$  is an estimator of the population probability  $\pi_I^+$  that driving is improved after using ZST. The  $m = N - (n_I^+ + n_I^-)$  subjects for whom  $-t^* \leq Y_I - Y_p \leq t^*$  are considered to have changes from one condition to the other that do not represent important modification of their driving skills.

The null hypothesis that driving after using ZST is impaired to the same degree as driving after using placebo can thus be formulated as

$$H_0: \pi_{I+} = \pi_{I-}$$

The null essentially declares symmetry of the underlying distribution of the SDLPs beyond the threshold values.

**Reviewer's Comment:** *This statistical test comparing the proportion impaired to the proportion "improved" may not always make clinical sense. In particular, if there is a well established SDLP cutoff corresponding to impairment and  $\pi_{I+} = \pi_{I-}$ , i.e., the proportion of impaired drivers equals the proportion of "improved" drivers, and  $\pi_{I+}$  is moderately large then even though the distribution is symmetric there is a moderately large proportion of impaired drivers. Ignoring this would be like ignoring a common serious adverse event which we would not do. Therefore, without further constraints on the size of the proportion impaired this symmetry test seems problematic from a regulatory perspective in this setting of driving safety. In other words, the sponsor's test may mathematically cancel out a serious clinical issue between two different groups, but the impaired group still exists. This is a cause for concern about the use of this test in this type of study.*

### **A statistic for testing the null hypotheses for impaired driving**

A one degree of freedom chi square statistic, the General Sign Test (GST), provides a test of the null hypothesis. The statistic is

$$(n_I^+ - n_I^-)^2 / (n_I^+ + n_I^-).$$

It has been shown (Cochran, 1937; Dixon and Mood, 1946) that McNemar's discordant matched pairs test, a well known and often used statistical procedure, is mathematically equivalent to a GST in which neutral observations are permitted. More generally, tests used in the matched pairs setting apply directly to the setting where the GST can be used and their statistical properties, including power (Hoover, 2005), are identical. Sahai and Khurshid (1996) reviewed the commonly used tests.

For example, for a given value of  $m$ , the distribution of  $n_I^+$  is Binomial with parameters  $m$  and  $\pi$  where

$$\pi = \pi_I^+ / (\pi_I^+ + \pi_I^-)$$

and  $\pi_I^+ = \pi_I^- = \pi = 1/2$  when  $H_0$  is true.

Therefore, an alternative approach to testing the null hypothesis is to condition on  $m$  and use an exact or large sample binomial test approximation.

The primary analysis is based on a binary indicator representing a difference in SDLP between active treatment and placebo that exceeded the threshold of 2.5 cm. As a sensitivity test of the dependence of the results on the choice of threshold value, two other cutpoints were defined as per the Clinical Study Protocol dated 21 May 2010, namely 2.0 and 3.5 cm. Moreover, because

there is no widely accepted consensus as to the threshold that distinguishes impaired from unimpaired drivers, as per the final SAP, a range of threshold values from 1.75 cm to 6.5 cm was evaluated in addition to the threshold levels of 2.0 cm and 3.5 cm that were specified in the secondary analyses in the Clinical Study Protocol.

All hypothesis tests of the primary outcome measure were pairwise contrasts using the McNemar's test. McNemar's test is equivalent to the GST. The first between treatment test was to compare ZST 3.5 mg with 4 hours of remaining sleep (Condition C) and placebo (Condition D). Next ZST with 3 hours of remaining sleep (Condition B) and placebo (Condition D) will be compared. Finally, zopiclone 7.5 mg (ZOP, Condition A) and placebo (Condition D) will be compared. It is expected that this null will be rejected to indicate that the study has sufficient sensitivity to detect differences in the desired effect range. Comparison of ZOP and the two ZST arms are of secondary interest but will be performed to obtain an estimate of their relative potential for producing impairment. If the first hypothesis test does not permit Conditions A1 and A2 to be combined, or condition D1 and D2 to be combined, tests will be performed within comparable design conditions and combined where possible. Secondary measures, SDLP and SDS, will be analyzed to compare mean effects using a repeated measures analysis of variance model for SDLP which will include fixed effects for sequence, period, and treatment, and a random effect for subject within sequence. The primary contrasts are between both ZST arms and placebo. Because this trial is designed to determine if there are negative effects produced by ZST compared to placebo, no adjustment for multiplicity will be made. However, this needs to be taken into account when interpreting results, for the possibility of finding spurious significant differences is substantially increased.

### **Determination of sample size**

With the threshold set at  $t^* = 2.5$ , a sample size of 36 is required to achieve .8 power to detect differences in symmetry between Probability (falling above  $t^*$ ) = .05 and Probability (falling below  $-t^*$ ) = .26. To achieve balance in the Williams designs, 40 subjects were to be enrolled. The test used is a Generalized Sign/McNemar's Test at a significance level 0.05. The power for .02 vs. .20 is .8; for .05 vs. .35 the power is .94; for .05 vs. .5 the power is .99. The assumptions underlying these calculations are based on pairwise comparisons of two studies of ZOP 7.5 mg vs. placebo (Table 9) (12/21/2009 and 12/23/2009 correspondence between Dr. Annemiek Vermeeren, principal investigator for both studies, and Transcept).

### **Initial Review of Driving Data**

During the driving test, the lateral position of the car relative to the left lane boundary and the car's speed are continuously recorded and digitally sampled at 4Hz.

Data was collected from the electronic source file and was reviewed to select/reject what were considered data artifacts. Examples of artifacts included: overtaking maneuvers, turn-around (half-way through drive), light reflections or darkness data errors, etc. Meetauto's Statistisch Programma (MSUP, the "Test Car's Statistics Program", version 1.6.0) was utilized for this purpose. The editing process was conducted by well-trained personnel who were blinded to the treatment assignments. It resulted in determination of a mean, minimum, maximum, median standard deviation of both the lateral position and speed. Two electronic files were created with extensions .SEG and .MRK. The .SEG file contains subject test scores and the .MRK file

contains the editing script (both available as source documents at the site). After all the subject data was reviewed, the .SEG files were utilized to prepare a single Excel file. In order to assure Excel file accuracy, the files were checked by additional blinded personnel. All these procedures are appropriately documented in the standard operating procedures of Maastricht University. These data are reduced by 10km segments of the ride and separate values are averaged over the entire ride to yield the overall mean, standard deviation (SD), the standard index of skew and that of kurtosis for both variables, separately.

*Reviewer's Comment: The first subject was enrolled on June 11, 2010 and the last one completed on 17 September, 2010. The standard operating procedures document for the data editing of the Highway Driving Test was signed on May 31, 2010 and the data editing forms were dated 04 May, 2010. Thus, it appears that the plan for removing the so-called "artifacts" was in fact prespecified. There is still some concern that there could be some misclassifications of genuinely impaired driving as artifacts but the editors were trained and there was some oversight of them.*

A driving test may be terminated prior to scheduled completion by decision of the driving instructor or the subject, if either determines that the subject is too drowsy to continue safely. Historically when this happens, it is usually between 50% and 100% completion of the driving test. Values of SDLP and SDS will be derived from the recorded driving data prior to termination. The driving data, which contains variables, such as subject ID, treatment period, SDLP, SDS, and if driving test was terminated early will be transferred to PRA in Excel format.

### **Secondary Analyses**

Additional secondary endpoints will be SDLP in centimeters in the highway driving test and standard deviation in speed (SDS). The analysis will be a repeated measures analysis of variance model that will include fixed effects for sequence, treatment period, and treatment, and a random effect for subject within sequence. The covariance structure will be assumed as compound symmetry in the ANOVA model. The estimation procedure will use REML. The analysis of ANOVA model will be performed on reported driving data (SDLP and SDS). Pairwise contrasts as described above will follow the ANOVA. The first two pairwise tests will compare ZOP Conditions A1 and A2 and Placebo Conditions D1 and D2 to determine if they can respectively be combined into Condition A (ZOP) and Condition D (placebo). If they could be combined, as the sponsor expected, the first between treatment contrast was to compare zolpidem tartrate sublingual tablet 3.5 mg (ZST) (Condition C) and placebo (Condition D).

#### **3.3.1.2 Baseline Demographics**

A total of 40 volunteers (20 female and 20 male) were enrolled and randomized to treatment schedules. Their median age (range) was 32 (21-64) years. Ethnic distribution was 0% Hispanic/Latino and 100% Not Hispanic or Latino. Race distribution was 97.5% white, 0% black, 0% Asian and 2.5% 'other'. The majority of study participants were non-smokers (87.5%), and reported to be consumers of alcohol (85.0%) and coffee (90.0%). There were no reports of relevant or clinically significant medical conditions or disorders in the medical history,

especially also no reports of insomnia or other sleep disorders. Table 2 shows the baseline demographics of the study population.

**Table 2 Baseline Demographics**

Variable	Class	n (%)	Mean	SD	Median	Min	Max
Age (years)		40	37	15	32	21	64
Height (cm)		40	176	9	178	158	195
Weight (kg)	Male	20	79	8	78	68	92
	Female	20	65	8	64	53	84
BMI (kg/m <sup>2</sup> )		40	23.2	2.4	23.1	18.5	27.9
Sex	Male	20 (50.0%)					
	Female	20 (50.0%)					
Ethnicity	Hispanic or Latino						
	Not Hispanic or Latino	40 (100%)					
Race	White	39 (97.5%)					
	Black	0					
	Asian	0					
	Other	1 (2.5%)					

Note: Table copied from page 51 of sponsor's study report

### 3.3.1.3 Randomly Assigned Crossover Sequences of Treatment

Subjects were randomized to 24 possible sequences for four treatments under the assumptions that A1 and A2 (Zopiclone at bedtime followed by placebo at 5:15 and 4:15, respectively), would be equivalent and D1 and D2 (placebo at bedtime followed by placebo at 5:15 and 4:15,



respectively) would be equivalent. The observed frequency of the various sequences of Treatments with A1,A2 and D1,D2 combined was as follows in Table 3.

**Table 3 Observed Frequency of Randomized Sequences of Treatment**

<i>Sequence</i>	<i>Frequency</i>
A-B-C-D	2
A-B-D-C	1
A-C-B-D	2
A-C-D-B	2
A-D-B-C	1
A-D-C-B	2
B-A-C-D	1
B-A-D-C	2
B-C-A-D	1
B-C-D-A	2
B-D-A-C	2
B-D-C-A	2
C-A-B-D	2
C-A-D-B	2
C-B-A-D	2
C-B-D-A	1
C-D-A-B	2
C-D-B-A	1
D-A-B-C	2
D-A-C-B	1
D-B-A-C	2
D-B-C-A	2
D-C-A-B	1
D-C-B-A	2

The observed frequency of various actual sequences of treatments distinguishing between A1 and A2 and D1 and D2 was as follows in Table 4.

**Table 4 Observed Frequency of Actual Sequences of Treatment**

<b>Sequence</b>	<b>Frequency</b>
A1/B/C/D2	1
A1/B/D2/C	1
A1/C/B/D2	1
A1/C/D1/B	1
A1/D2/B/C	1
A2/B/C/D1	1
A2/C/B/D1	1
A2/C/D1/B	1
A2/D1/C/B	1
A2/D2/C/B	1
B/A1/C/D1	1
B/A1/D2/C	1
B/A2/D1/C	1
B/C/A1/D1	1
B/C/D2/A2	2
B/D1/A2/C	1
B/D1/C/A1	1
B/D2/A2/C	1
B/D2/C/A2	1
C/A1/D2/B	1
C/A2/B/D2	2
C/A2/D1/B	1
C/B/A1/D1	1
C/B/A2/D2	1
C/B/D1/A1	1
C/D1/A1/B	1
C/D1/A2/B	1
C/D1/B/A1	1
D1/A1/C/B	1
D1/A2/B/C	1
D1/B/C/A2	1
D1/C/A1/B	1
D1/C/B/A2	1
D2/A2/B/C	1
D2/B/A1/C	1
D2/B/A2/C	1
D2/B/C/A1	1
D2/C/B/A1	1

### 3.3.1.4 Subject Disposition

The protocol defines the ITT analysis set as all randomized subjects who received at least one dose of study drug. In this plan, the ITT analysis set is defined as all randomized subjects who received at least one dose of study drug and commenced at least one drive. Subjects without any driving data were not to be included in the primary and secondary analyses. All 40 subjects received study drug in all 4 periods and underwent all 160 driving tests (40 tests per period).

#### Prematurely terminated driving tests

A total of 160 driving tests were performed by 40 subjects. Three (3) driving tests were terminated prematurely in 2 subjects, because the driving instructor considered the subject to be too drowsy to continue safely. Two (2) driving tests were stopped prematurely after ZOP and 1 test was stopped prematurely after ZST 3h at 70 min, 43 min and 44 min, respectively, after the start of the driving test.

### 3.3.1.5 Results

With four treatments there are 24 possible ordered sequences for administering the treatments to a patient and with 6 treatments there are 720. In this William's Design each treatment, if we consider A1 and A2 the same and D1 and D2 the same, occurs 10 times in each of the four positions.

A shortcoming of the design utilized is that it assumes that placebo administered at 5:15 or 4:15 has the same effect. With this design there is no way to test this using each patient as his/her own control since each patient only got assigned to one or the other. Although it's not as ideal as having each subject be their own control (e.g., less power) we can compare those assigned to D1 with those assigned to D2. Given these limitations, there don't appear to be any significant differences in SDLPs between groups of patients assigned to the different placebos, denoted D1 and D2, (Wilcoxon,  $p=.2387$ ) during the placebo period. Therefore, we can use this as some justification for treating D1 and D2 as the same placebo in subsequent analyses. Table 5 summarizes the distribution of SDLP for each placebo condition.

**Table 5 Summary Statistics of SDLP for the Two Placebo Treatments**

<i>TRT</i>	<i>N</i>	<i>MEAN</i>	<i>S.D.</i>	<i>MIN</i>	<i>25%ILE</i>	<i>MEDIAN</i>	<i>75%ILE</i>	<i>MAX</i>
D1(3hrs)	21	15.36	3.52	8.70	12.83	16.33	16.66	22.17
D2(4hrs)	19	16.45	2.62	11.64	14.20	17.05	18.59	20.24

Table 6 shows the sponsor's results for the active control ZOP (note that conditions D1 and D2 were combined and A1 and A2 were combined). The achievement of nominal significance for the active control at multiple thresholds suggests that the study had assay sensitivity.

**Table 6 Number Impaired SDLP Changes of Active Control from Placebo**

Treatment Versus Placebo	Number of Subjects				Probability		McNemar Statistic	p-value
	Threshold (cm)	Impaired	Neutral	Improved	Impaired	Improved		
ZOP	1.75	22	18	0	0.550	<.001	22.0	<.0001
	2	19	21	0	0.475	<.001	19.0	<.0001
	2.25	19	21	0	0.475	<.001	19.0	<.0001
	2.5	18	22	0	0.450	<.001	18.0	<.0001
	2.75	17	23	0	0.425	<.001	17.0	<.0001
	3	16	24	0	0.400	<.001	16.0	<.0001
	3.25	16	24	0	0.400	<.001	16.0	<.0001
	3.5	14	26	0	0.350	<.001	14.0	0.0001
	3.75	13	27	0	0.325	<.001	13.0	0.0002
	4	13	27	0	0.325	<.001	13.0	0.0002
	4.25	11	29	0	0.275	<.001	11.0	0.0010
	4.5	10	30	0	0.250	<.001	10.0	0.0020
	4.75	9	31	0	0.225	<.001	9.00	0.0039
	5	9	31	0	0.225	<.001	9.00	0.0039
	5.25	8	32	0	0.200	<.001	8.00	0.0078
	5.5	6	34	0	0.150	<.001	6.00	0.0313
	5.75	5	35	0	0.125	<.001	5.00	0.0625
	6	4	36	0	0.100	<.001	4.00	0.1250
	6.25	4	36	0	0.100	<.001	4.00	0.1250
	6.5	4	36	0	0.100	<.001	4.00	0.1250

Table 7 shows the sponsor's results for the difference in SDLP between ZST 4 hour and Placebo (D1 and D2 pooled).

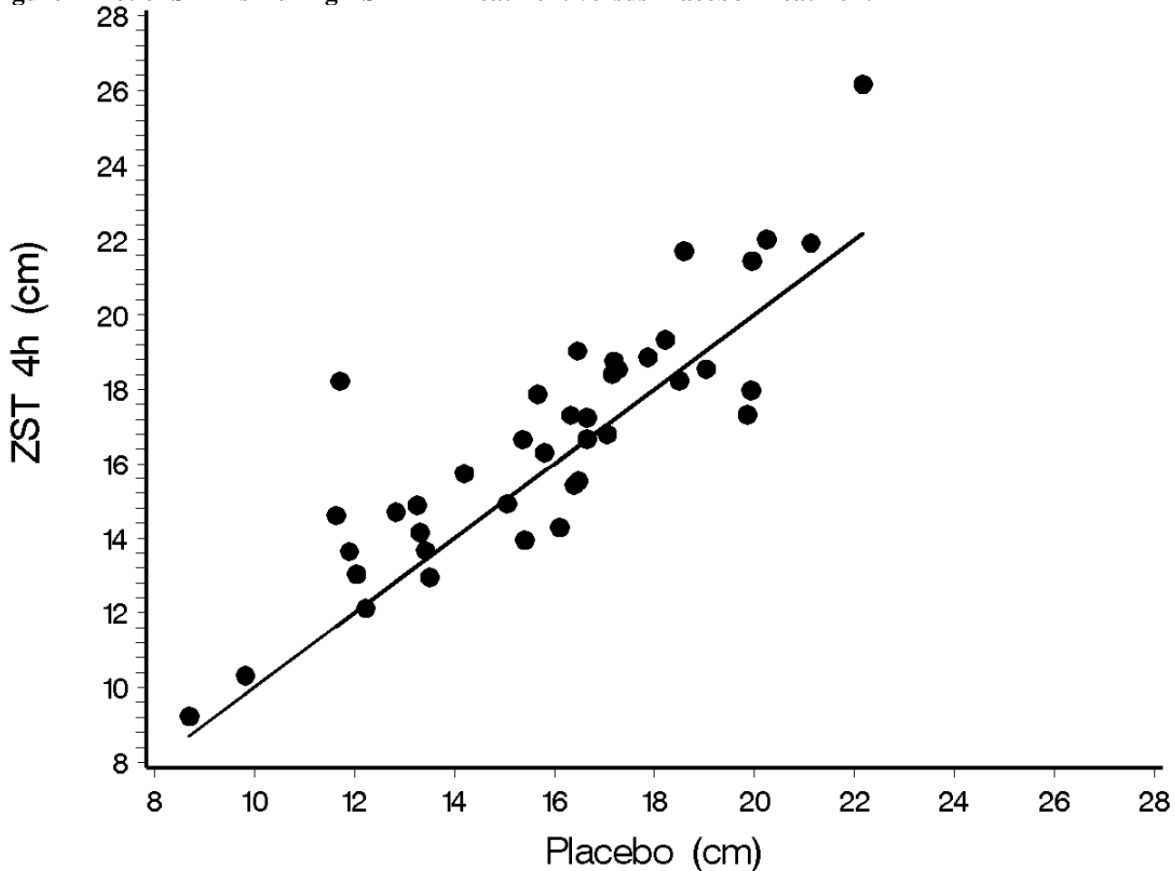
**Table 7 Number of ZST 4hr Changes from Placebo exceeding Various SDLP Cutpoints**

Treatment Versus Placebo	Number of Subjects				Probability		McNemar Statistic	p-value
	Threshold (cm)	Impaired	Neutral	Improved	Impaired	Improved		
ZST 4h	1.75	8	29	3	0.200	0.075	2.27	0.2266
	2	6	33	1	0.150	0.025	3.57	0.1250
	2.25	5	34	1	0.125	0.025	2.67	0.2188
	2.5	5	34	1	0.125	0.025	2.67	0.2188
	2.75	4	36	0	0.100	<.001	4.00	0.1250
	3	3	37	0	0.075	<.001	3.00	0.2500
	3.25	2	38	0	0.050	<.001	2.00	0.5000
	3.5	2	38	0	0.050	<.001	2.00	0.5000
	3.75	2	38	0	0.050	<.001	2.00	0.5000
	4	1	39	0	0.025	<.001	1.00	1.0000
	4.25	1	39	0	0.025	<.001	1.00	1.0000
	4.5	1	39	0	0.025	<.001	1.00	1.0000
	4.75	1	39	0	0.025	<.001	1.00	1.0000
	5	1	39	0	0.025	<.001	1.00	1.0000
	5.25	1	39	0	0.025	<.001	1.00	1.0000
	5.5	1	39	0	0.025	<.001	1.00	1.0000
	5.75	1	39	0	0.025	<.001	1.00	1.0000
	6	1	39	0	0.025	<.001	1.00	1.0000
	6.25	1	39	0	0.025	<.001	1.00	1.0000
	6.5	0	40	0	<.001	<.001		

Note: copied from page 53 of sponsor's study report

There were no statistically significant treatment effects at any threshold examined by the sponsor. Thus, there is not enough evidence of asymmetry of the underlying distribution of the SDLPs based on these thresholds. Figure 2 shows a scatter plot of the SDLP (cm) driving results for ZST 4h versus placebo for each subject. In this figure, the diagonal reference line is the reference line of equality of SDLP results for ZST 4h and placebo. Differences are distributed more or less symmetrically around the reference line although a majority do fall above the line.

**Figure 1 Plot of SDLPs During ZST 4hr Treatment versus Placebo Treatment**



Note: the reference line represents equal SDLP values for ZST 4h and Placebo

Note: copied from page 54 of sponsor's study report

A Non-parametric rank test is less susceptible to the influence of outliers than a typical analysis of means since the actual value is replaced in the analysis by an integer from 1 to 40, in this case (assuming no tied ranks), depending on the rank of the observed value. The sign test discards the actual value altogether except for the sign of the value. The sign test is actually equivalent to the sponsor's analysis method using a cutoff, but with a cutoff of 0, which they did not specify. The Signed rank test is more powerful than the sign test because it incorporates the sign of the observed value as well as its rank. Here, the signed rank test has a p-value of 0.0019. In this case even the less powerful Sign test suggests that changes from placebo in SDLP for ZST 4 hour have a significant tendency to be positive rather than negative,  $p=0.0166$ .

This reviewer also found that there are some other SDLP cutpoints not examined by the sponsor for which ZST 4 hour changes from placebo were nominally significant in a manner suggesting drug impairment of driving caused by ZST 4 hour. For a cutpoint of 2.55 there were 5 impaired vs. 0 improved, suggestive frequencies, but above nominal significance,  $p=0.0625$ . However nominal significance was observed for cutpoints of 1.5 and 1.475:

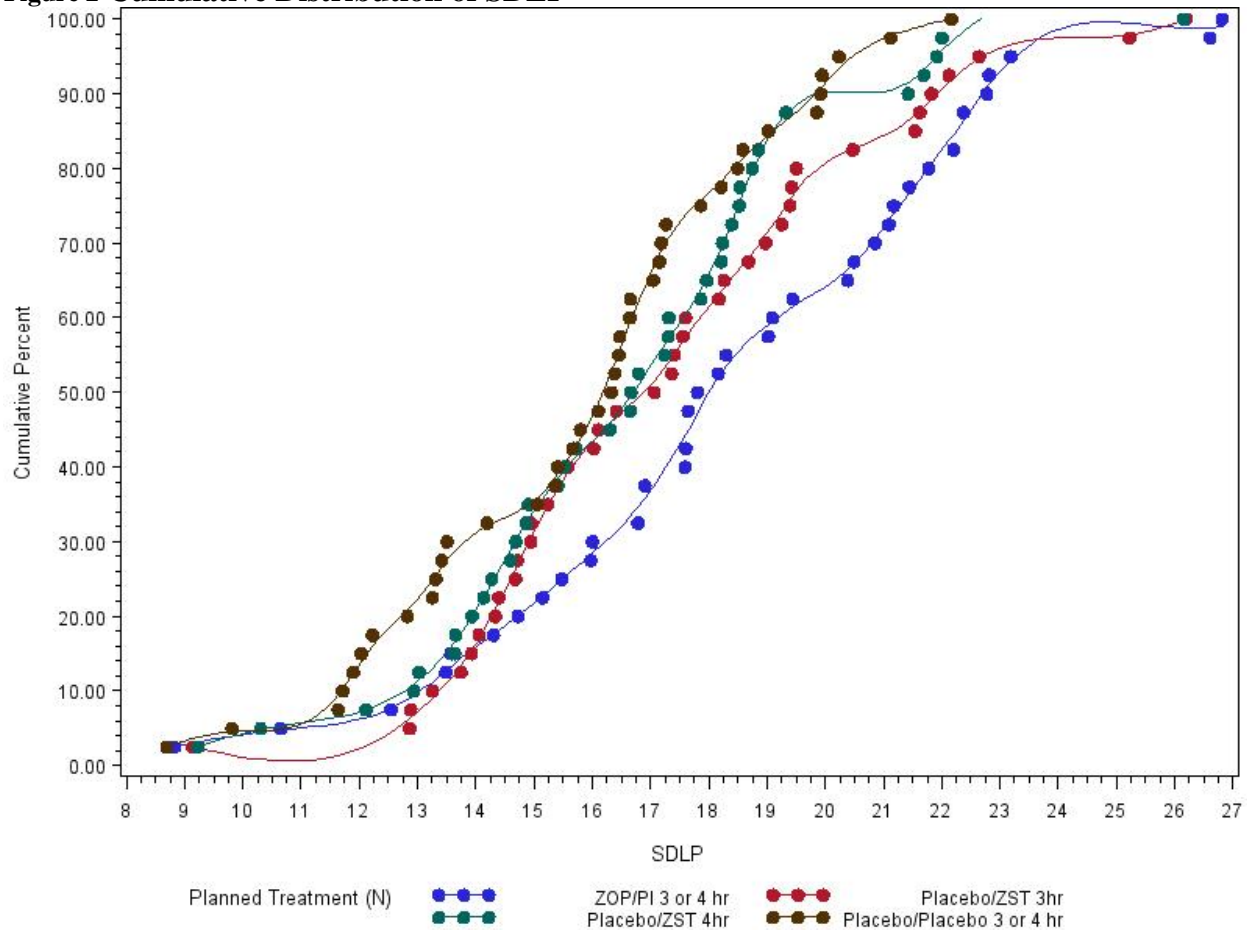
$p=0.03516$  with cutpoint of 1.500 (12 impaired vs. 3 improved)

$p=0.02127$  with cutpoint of 1.475 (13 impaired vs. 3 improved).

Two (2) active control subjects and 1 subject in ZST 3 hr fell asleep (a total of 2 subjects for these 3 events) while performing the driving test and thus had that test terminated. The subject that fell asleep during ZST 3 hour, subject 0010 a 23 year old female, had an increase in SDLP over placebo of 3.45. This obviously might have been more if the test was not terminated early.

The following figure shows the cumulative distribution of SDLP for each treatment. We need to be cautious in overinterpreting this since each patient received all four treatments at various times which is not typically the case for such graphs.

**Figure 2 Cumulative Distribution of SDLP**



The sponsor's table of changes in SDLP of ZST 3hr from placebo (Table 8 below) shows nominal significance for numerous cutpoints. Thus, it appears that taking ZST in the middle of the night 3 hours before driving leads to impaired driving.

**Table 8 Number of ZST 3hr Changes from Placebo exceeding Various SDLP Cutpoints**

Treatment Versus Placebo	Number of Subjects				Probability		McNemar Statistic	p-value
	Threshold (cm)	Impaired	Neutral	Improved	Impaired	Improved		
ZST 3h	1.75	16	21	3	0.400	0.075	8.89	0.0044
	2	13	25	2	0.325	0.050	8.07	0.0074
	2.25	11	27	2	0.275	0.050	6.23	0.0225
	2.5	10	29	1	0.250	0.025	7.36	0.0117
	2.75	10	29	1	0.250	0.025	7.36	0.0117
	3	10	30	0	0.250	<.001	10.0	0.0020
	3.25	8	32	0	0.200	<.001	8.00	0.0078
	3.5	7	33	0	0.175	<.001	7.00	0.0156
	3.75	6	34	0	0.150	<.001	6.00	0.0313
	4	6	34	0	0.150	<.001	6.00	0.0313
	4.25	4	36	0	0.100	<.001	4.00	0.1250
	4.5	3	37	0	0.075	<.001	3.00	0.2500
	4.75	3	37	0	0.075	<.001	3.00	0.2500
	5	2	38	0	0.050	<.001	2.00	0.5000
	5.25	2	38	0	0.050	<.001	2.00	0.5000
	5.5	2	38	0	0.050	<.001	2.00	0.5000
	5.75	2	38	0	0.050	<.001	2.00	0.5000
	6	2	38	0	0.050	<.001	2.00	0.5000
	6.25	1	39	0	0.025	<.001	1.00	1.0000
	6.5	1	39	0	0.025	<.001	1.00	1.0000

Note: copied from page 56 of sponsor's study report



### Secondary Analysis of SDLP as a Continuous Variable (No cutoff/threshold)

Both ZST 4hour and ZST 3hour were nominally significant compared to placebo in terms of mean SDLP as determined from the sponsor's prespecified secondary analysis, a repeated measures analysis of variance.

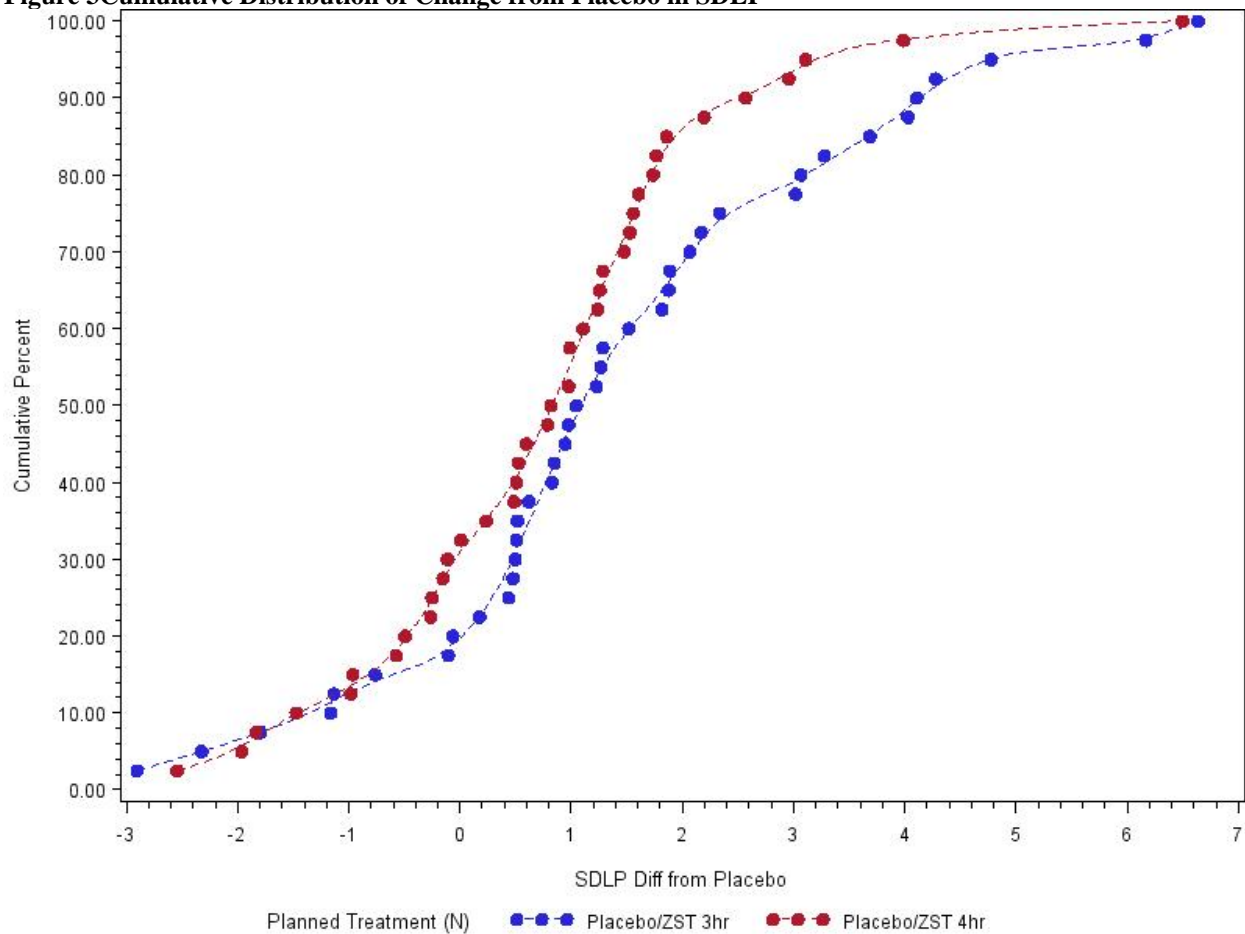
**Table 9 Secondary Analysis of SDLP as a Continuous Variable (No cutpoint)**

	DIFF	S.E.	95% LL	95% UL	P-VALUE
ZST 4hr- Placebo	0.8290	0.3434	0.1487	1.5093	.0174
ZST 3hr- Placebo	1.4560	0.3434	0.7757	2.1363	<.0001

The active control was also nominally significant with an estimated difference from placebo of 2.4577 +/- .3434 (S.E.),  $p < 0.0001$ . An exploratory comparison of ZST 3hr vs ZST 4hr based on the same model yielded an estimated difference of 0.627,  $p = 0.0705$ .

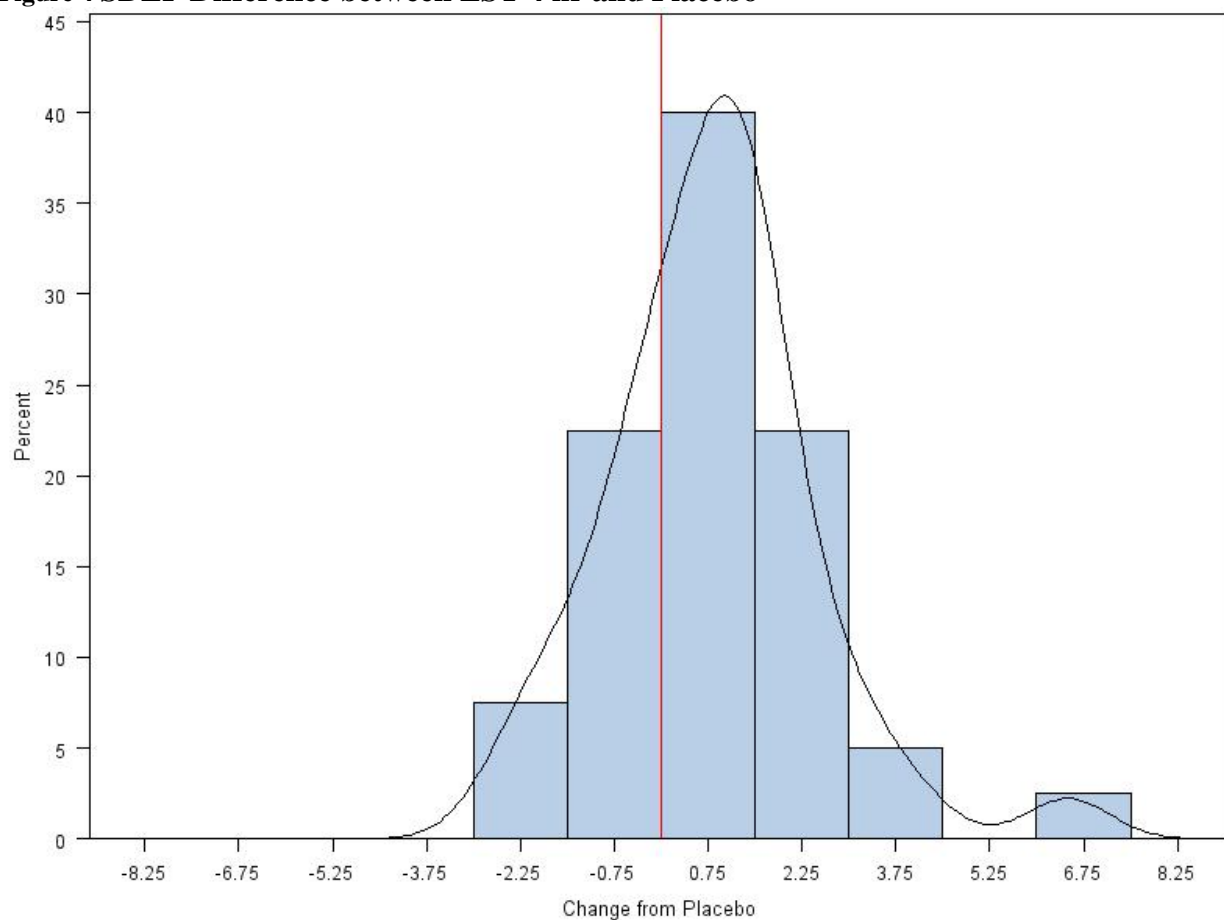
Note that a nominally significant period effect was found in this analysis  $p = 0.0171$ . A first period only analysis would have only 10 patients per group. This was nominally significant for ZOP ( $p = 0.0138$ ) but not for the ZST groups possibly due to loss of power from excluding periods 2-4. Figure 3 provides a graphical comparison of the distribution of differences from placebo in SDLP for ZST 3 hour and ZST 4 hour.

**Figure 3 Cumulative Distribution of Change from Placebo in SDLP**



The following histogram of differences between ZST 4 hr and Placebo suggests that the distribution of such differences is shifted slightly to the right of zero.

**Figure 4 SDLP Difference between ZST 4 hr and Placebo**



### **Exploratory Comparisons between ZST 3hour and ZST 4 hour**

Table 10 presents exploratory analyses of differences in SDLP between ZST 3 hour and ZST 4 hour using the test of symmetry. Note that for each row the total number of patients is 40 but the number of subjects between the upper and lower cutoff, 40-NImproved-NImpaired, is not shown because McNemar's test only depends on the two numbers (impaired or improved) shown.

Table 10 Differences in SDLP between ZST 3hr and ZST 4 hr

CUTOFF	N IMPROVED	N IMPAIRED	TEST OF SYMMETRY P-VALUE
1.50	5	13	.0963
1.75	3	10	.0923
2.00	3	8	.2266
2.25	3	7	.3438
2.50	2	6	.2891
3.00	2	6	.2891
3.50	1	4	.3125
3.75	1	2	1.0
4.00	1	1	1.0

The sign test (corresponding to a cutoff of 0) p-value for the comparison of these two groups is .0807 and signed rank test p is .0380. Although this last test is nominally significant none of the other cutoffs are and, anyways, the larger issue may be that ZST 4 hour was nominally significant compared to placebo for some cutoffs.

### 3.3.1.6 Reviewer's Analysis of All Driving Data Including "Artifacts"

If we keep the "data artifacts" in the analysis dataset then for the 2.5 cutoff of the difference from placebo we find the following.

Table 11 Differences from placebo in SDLP with all of raw data, i.e., no artifacts removed

TREATMENT	N IMPROVED	N IMPAIRED	TEST OF SYMMETRY P-VALUE
ZOP	2	13	p=0.0074
ZST 4 hour	8	10	p=0.8145
ZST 3 hour	5	9	p=0.4240

The signal for ZST 3hr that was observed in the sponsor's analysis may be lost in all the noise associated with passing slow cars, etc. It seems that an established procedure was in place for removing lane changes and similar things permitted during the test which may appear as impaired driving but are, in fact, not. Therefore, the study would have been powered under the assumption that artifacts would be removed before analysis. Although excluding data always arouses suspicions it was prespecified in this case for legitimate reasons associated with conducting the driving test on the open road. Therefore, it seems justified in this case and we will rely on the "artifact" free data.

At least for the ZST 3 hour treatment, because the difference from placebo was nominally significant for several of the prespecified cutpoints we need not worry that the removal of data “artifacts” obscured a driving impairment signal for that comparison.

### 3.3.1.7 Additional Reviewer Comments on the SDLP as a measure of impaired driving

Note that the SDLP is a summary of the entire driving test and in summarizing we lose the ability to infer anything about differences during shorter parts of the test. A certain difference in SDLPs between a patient’s test under ZST 3hour treatment and test under placebo could have various explanations. For example, it could be that there was a fairly consistent but very small difference in lateral position over the whole test or alternatively, it could be that for most of the test there were no differences in lateral position but there was one short epoch in which the subject went out of the lane under drug treatment. The SDLP would not enable us to differentiate between these two scenarios without looking at the raw continuous driving data from which the standard deviation of lateral position was determined. In summary, the SDLP doesn’t necessarily tell us how many impaired epochs of driving a subject had during the test on drug treatment as compared to placebo treatment.

It seems to this reviewer that there may be a shortcoming of the SDLP for assessing driving safety. In particular, outliers may yield a mean estimate which is different from the lateral position corresponding to the car being in the exact middle of the lane position and the SDLP measures the deviations from the mean but it seems like one might be more interested in deviations from the middle of the lane on the road. It seems like a different mean than the value corresponding to the middle of the lane would be a concern in itself. Among two individuals with the same  $\sigma$  one of whom had a non-middle mean the one with the non-middle mean would tend to have slightly larger deviations from the middle. This can be seen from the following equation. Suppose  $\mu$  is the lateral position measured by the camera when the car is in the exact middle of the intended driving lane.

$$\sigma_{\mu}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y} - [\bar{y} - \mu])^2 = \sigma_{\bar{y}}^2 + \frac{n}{n-1} (\bar{y} - \mu)^2$$

Note that in this equation the cross product term cancels because  $\bar{y} - \mu$  is constant in  $i$  and  $y_i - \bar{y}$  sums to  $n\bar{y} - n\bar{y} = 0$ .

Other researchers have suggested using the occurrence of going out of the lane as the outcome but it might be too rare of an event to be practical. Another option instead of  $\sqrt{\sigma_{\bar{y}}^2}$  is  $\sqrt{\sigma_{\mu}^2}$ .

Of course, even if there was a signal on this new measure, it would not be definitive because of the exploratory nature of this alternative measure.

## 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

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

#### 4.1.1 Gender

Fifty percent of the 40 study subjects were male and 50% were female. In males 3 out of 20 had an increase in SDLP exceeding 2.5 and 1 had a decrease below -2.5.

In females 2 out of 20 had an increase in SDLP exceeding 2.5 and 0 had a decrease below -2.5. While there may be a hint in the following data of a greater safety issue for females than males this study is too small as to be conclusive in this regard.

Table 12 shows gender subgroup specific tests of symmetry of the distribution of differences of ZST 3 hour from placebo. Note that for each row the total number of patients is 40 but the number of subjects between the upper and lower cutoff, 40-NImproved-NImpaired, is not shown because McNemar's test of symmetry only depends on the number impaired and the number improved.

**Table 12 Differences between ZST 3hour and Placebo Exceeding Various Thresholds by Gender**

GENDER	CUTOFF	IMPROVED	IMPAIRED	TEST OF SYMMETRY PVALUE
M	1.75	2	7	.1797
M	2.0	1	5	.2188
M	2.25	1	5	.2188
M	2.5	0	4	.1250
M	2.75	0	4	.1250
M	3.0	0	4	.1250
F	1.75	1	9	.0215
F	2.0	1	8	.0391
F	2.25	1	6	.1250
F	2.5	1	6	.1250
F	2.75	1	6	.1250
F	3.0	0	6	.0313

In females 16 of 20 (80%) had numerically higher SDLP on ZST 4 hr than on placebo as compared to 12 of 20 (60%) in males. Table 13 shows gender subgroup specific tests of symmetry of the distribution of differences in SDLP between ZST 4 hour and Placebo.

**Table 13 Differences between ZST 4hour and Placebo Exceeding Various Thresholds by Gender**

GENDER	CUTOFF	IMPROVED	IMPAIRED	TEST OF SYMMETRY PVALUE
M	1.5	3	4	1.0
M	1.75	3	4	1.0
M	2.0	1	4	.375
M	2.25	1	3	.625
M	2.5	1	3	.625
M	2.75	0	3	.25
M	3	0	2	.50
F*	1.5	0	8	.0078
F	1.75	0	4	.125
F	2.0	0	2	.50
F	2.25	0	2	.50
F	2.5	0	2	.50
F	2.75	0	1	1.0
F	3.0	0	1	1.0

\*cutoff not specified by sponsor

### 4.1.2 Race

Only one of the 40 subjects was non-White so it is not possible to discern whether there are differential drug effects by Race in this study.

### 4.1.3 Age

Ages ranged from 21 to 64. Therefore, no analysis comparing ages < 65 to ages ≥65 is possible. The mean age was 37.3 and the median age was 31.5.

For the ZST 3 hour there might be an age effect since those below the median age had significantly more beyond various cutoffs of differences from placebo in SDLP. However, there was at least one cutoff for which ages above the median showed impairment as well. Table 14 shows age subgroup specific tests of symmetry of the distribution of differences in SDLP between ZST 3 hour and Placebo.

**Table 14 Differences between ZST 3hour and Placebo Exceeding Various Thresholds by Age Group**

AGE GROUP	CUTOFF	IMPROVED	IMPAIRED	TEST OF SYMMETRY PVALUE
Age <31.5	1.75	2	12	.0129
	2.00	1	10	.0117
	2.25	1	9	.0215
	2.50	0	8	.0078
Age ≥31.5	1	1	8	.039
	1.75	1	4	.375
	2	1	3	.625
	3	0	2	.500

Table 15 shows age subgroup specific tests of symmetry of the distribution of differences in SDLP between ZST 4 hour and Placebo. There are no obvious differences between the age groups for ZST 4hour and placebo.

**Table 15 Differences between ZST 4hour and Placebo Exceeding Various Thresholds by Age Group**

AGE GROUP	CUTOFF	IMPROVED	IMPAIRED	TEST OF SYMMETRY PVALUE
Age < 31.5	1.75	2	6	.289
	2.00	0	4	.125
	2.25	0	3	.250
	2.50	0	3	.250
Age ≥ 31.5	1.50	1	4	.375
	2.00	1	2	1.00
	2.50	1	2	1.00
	3.00	0	1	1.00



As seen in Table 16 there doesn't appear to be any differential effect of age above or below the median for the active control.

**Table 16 Differences between ZOP (Active Control) and Placebo Exceeding Various Thresholds by Age Group**

AGE GROUP	CUTOFF	IMPROVED	IMPAIRED	TEST OF SYMMETRY PVALUE
Age<31.5	2	0	10	.0020
	3	0	8	.0078
Age ≥ 31.5	2	0	9	.0039
	3	0	8	.0078

#### **4.1.4 US vs. non-US**

This study was conducted entirely outside the US, so no comparison of US versus non-US driving safety after middle of the night treatment with ZST is possible on the basis of this study.

#### **4.2 Other Special/Subgroup Populations**

No other subgroups were analyzed.

## **5. SUMMARY AND CONCLUSIONS**

### **5.1 Statistical Issues and Collective Evidence**

Efficacy of treating middle of the night awakening was evaluated in a previous submission. This submission focused on next day driving safety. A single crossover designed study of next day driving safety after treatment with Zolpidem sublingual demonstrated assay sensitivity by means of the comparison of the active control, Zopiclone, to placebo. The primary endpoint was the standard deviation of lateral position on the road as obtained from a standardized driving test. The primary analysis involved the number of driving impaired subjects, defined as those having an increase from placebo in SDLP under drug treatment that exceeded 2.5 cm. There is some question about the most relevant cutoff of the distribution of differences from placebo in SDLP used to define impairment so the sponsor also prespecified several other cutoffs. Because this is a

safety study there was no multiplicity adjustment applied for examining more than one cutoff. In this analysis the number of impaired subjects was compared to the number of improved subjects, defined as those having a decrease from placebo in SDLP under drug treatment that was below - 2.5 cm. The SDLP data suggests that ZST 3 hour, i.e., taking ZST 3 hours before awakening, impaired next day driving. Although ZST 4 hour did not reach nominal significance for the primary cutoff of 2.5 cm, or several other cutoffs specified by the sponsor, there were still other cutoffs in addition to the prespecified secondary analysis of mean differences which suggested impairment for ZST 4 hour compared to placebo. There is a multiplicity issue related to using several cutoffs but we are dealing with a potentially very serious safety issue and a questionable cutoff designated as primary. Thus, overall, it seems to this reviewer that there may be some impairment of next day driving caused by taking ZST 4 hours before rising as well.

This study utilized a four period crossover design. Two dosing times, 3 hours and 4 hours before rising time, were investigated but the sponsor decided that it was not practical or necessary to have each subject undergo the six periods of scientific interest (investigational 3 and 4 hour, placebo 3 and 4 hour, and active control 3 hour and 4 hour). Therefore, all subjects underwent both 3 hour and 4 hour for ZST in different periods but they underwent only one time for placebo and only one time for active control. If it turned out that placebo 3 hours before rising and placebo 4 hours before rising had different effects on next day driving then the study may have been underpowered.

There was a prespecified procedure for having trained, blinded personnel edit the continuously sampled driving data prior to analysis in order to remove anticipated data “artifacts” of the driving test, such as passing slow cars, not enough light entering the camera, etc. Although excluding data always arouses suspicions it was prespecified in this case for legitimate reasons associated with conducting the driving test on the open road and, ostensibly, proper blinding of the editors was used. Of course, there is some risk of misclassification. However, it is important to distinguish between truly impaired driving and permitted driving behavior that could appear as impaired, such as passing a slow car. Therefore, it seems justified in this case and we will rely on the “artifact” free driving data. This would probably have been more of a concern if the investigational drug showed no evidence of impairment in the edited data but showed evidence in the unedited data. Both forms of the data were examined and this was found not to be the case.

Single site studies such as this may reduce variability but may have less generalizability.

## **5.2 Conclusions and Recommendations**

The SDLP crossover study data suggests that ZST 3 hour, i.e., taking ZST 3 hours before awakening, impaired next day driving. Although ZST 4 hour did not reach nominal significance for the primary cutoff of 2.5 cm, or several other cutoffs specified by the sponsor, there were still other cutoffs in addition to the prespecified secondary analysis of mean differences which suggested impairment for ZST 4 hour compared to placebo. There is a multiplicity issue related to using several cutoffs but we are dealing with a potentially very serious safety issue and a questionable cutoff designated as primary. Thus, overall, it seems to this reviewer that there may be some impairment of next day driving caused by taking ZST 4 hours before rising as well.

-----  
**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  
07/07/2011

KUN JIN  
07/07/2011  
I concur with this ewview.

HSIEN MING J J HUNG  
07/08/2011



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:** 22,328

**Drug Name:** Zolpidem Tartrate

**Indication(s):** Insomnia: Middle of the Night Awakenings

**Applicant:** Transcept

**Date(s):** September 30, 2008

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

**Clinical Reviewer:** Carole Davis, M.D.

**Project Manager:** Cathleen Michaloski

**Keywords:**  
Crossover Design, Multiplicity

# Table of Contents

<b>LIST OF TABLES.....</b>	<b>3</b>
<b>LIST OF FIGURES.....</b>	<b>4</b>
<b>1 EXECUTIVE SUMMARY .....</b>	<b>5</b>
1.1 CONCLUSIONS AND RECOMMENDATIONS .....	5
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES .....	5
1.3 STATISTICAL ISSUES AND FINDINGS .....	5
<b>2 INTRODUCTION .....</b>	<b>6</b>
2.1 OVERVIEW.....	6
2.2 DATA SOURCES .....	7
<b>3 STATISTICAL EVALUATION .....</b>	<b>7</b>
3.1 EVALUATION OF EFFICACY .....	7
3.1.1 Study ZI-06-010 .....	7
3.1.1.1 Study Plan .....	7
3.1.1.2 Demographics and Baseline Characteristics.....	12
3.1.1.3 Patient Disposition .....	15
3.1.1.4 Sponsor's Results.....	16
3.1.1.5 Reviewer's Results.....	18
3.1.1.5.1 Primary Analysis.....	18
3.1.1.5.2 Analysis of Crossover Period 1 Only .....	18
3.1.1.5.3 Investigation of the Impact of Missing Data on the Primary Analysis Result.....	18
3.1.1.5.4 Comparison of LPS pre MOTN awakening across Groups and Crossover Periods .....	19
3.1.1.5.5 Analyses of Secondary Endpoints .....	23
3.1.2 Study ZI-12.....	24
3.1.2.1 Overall Study Design and Plan .....	24
3.1.2.2 Demographic and Other Baseline Characteristics .....	27
3.1.2.3 Patient Disposition .....	28
3.1.2.4 Sponsor's Results.....	29
3.1.2.5 Reviewer's Results.....	38
3.1.2.5.1 Primary Analysis.....	38
3.1.2.5.2 Assessment of the Impact of Missing Data on the Primary Analysis Result.....	38
3.1.2.5.3 Analyses of Key Secondary Endpoints .....	38
3.1.2.5.4 Observed Treatment Effects by Site.....	39
3.2 EVALUATION OF SAFETY .....	40
<b>4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>41</b>
4.1 GENDER, RACE AND AGE .....	41
4.1.1 Gender .....	41
4.1.2 Race .....	41
4.1.3 Age.....	42
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS .....	42
<b>5 SUMMARY AND CONCLUSIONS .....</b>	<b>43</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	43
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	44

## LIST OF TABLES

Table 1 Study ZI-06-010: Demographics and Baseline Characteristics .....	13
Table 2 Study ZI-06-010: Patient Disposition .....	15
Table 3 Study ZI-06-010 Dosing Errors .....	16
Table 4 Study ZI-06-010: Sponsor's Results for LPS Post MOTN Awakening .....	17
Table 5 Study ZI-06-010: LPS before Scheduled Awakening .....	19
Table 6 Study ZI-06-010: LPS before Scheduled Awakening Day 1 of Crossover Period .....	21
Table 7 Study ZI-06-010: LPS before Scheduled Awakening Day 2 of Crossover Period .....	22
Table 8 Study ZI-12: Baseline Demographics .....	28
Table 9 Study ZI-12: Latency to Sleep Onset Post Middle-of-the-Night Awakening after Dosing .....	30
Table 10 Study ZI-12: Sponsor's Analysis of Subjective Total Sleep Time .....	32
Table 11 Study ZI-12: Sponsor's Analysis of Number of Awakenings after Treatment .....	35
Table 12 Study ZI-12: Sponsor's Post-Hoc Analysis of Subjective WASO .....	37

**LIST OF FIGURES**

Figure 1 Study ZI-06-010: LPS before MOTN awakening (before drug administration) .....	20
Figure 2 Study ZI-12: Disposition of Subjects .....	29
Figure 3 Study ZI-12 Treatment Effect on Log(Latency to Sleep Onset) by Site .....	40

## **1 EXECUTIVE SUMMARY**

### **1.1 Conclusions and Recommendations**

The data from the key efficacy studies, ZI-06-010 and ZI-12, seem to support the superiority of the Zolpidem tartrate sublingual lozenge over placebo for reducing latency to persistent sleep after a middle of the night awakening.

### **1.2 Brief Overview of Clinical Studies**

Two clinical studies are presented to support the efficacy of Zolpidem tartrate sublingual lozenge for Middle of the night awakenings associated with Insomnia in the NDA. Study ZI-06-010 was a Randomized, Double-blind, Placebo-controlled, Crossover Study of the Efficacy and Safety of the Sublingual zolpidem tartrate lozenge in Adult Patients with Insomnia Characterized by Difficulty Returning to Sleep after Middle-of-the-Night (MOTN) Awakening.

The other study, ZI-12, was a 4-week double-blind, placebo-controlled, self-report outpatient study, which enrolled patients with primary insomnia characterized by spontaneous middle-of-the-night awakenings with difficulty returning to sleep. Both studies were conducted within the United States.

### **1.3 Statistical Issues and Findings**

In study ZI-06-010 the Latency to Persistent Sleep before scheduled MOTN awakening was measured each study night. There was some suggestion in the data that the Latency to Persistent Sleep (LPS) pre-MOTN awakening varied significantly with the crossover period as well as the treatment group, despite it corresponding to the period before drug administration. However, it was prespecified that the primary analysis of the primary endpoint, LPS post MOTN awakening, would use the Total Sleep Time (TST) before MOTN awakening as the covariate instead of using the LPS pre MOTN awakening as the covariate. Furthermore, there was no period or treatment effect apparent on the TST before MOTN awakening. The primary analysis found that Zolpidem 3.5 mg was superior to placebo in reducing Latency to Persistent Sleep as measured by PSG. In spite of this possible group difference in LPS prior to scheduled awakening the pre-specified primary analysis as well as most other analyses (e.g., restricted to the first of two days in each crossover period or restricted to a particular crossover period) seem to support a significant treatment effect on LPS after scheduled awakening. Therefore, in light of the balance of the data suggesting a robustness of the statistical significance of the treatment effect we may be inclined to disregard the observed group differences in pre-MOTN LPS. Note again also that the primary analysis of LPS post MOTN prespecified adjusting for total sleep time before scheduled awakening rather than LPS before scheduled awakening and there were no significant group imbalances in total sleep time before scheduled awakening.



The analysis plan stated that following secondary endpoints based on the time after the scheduled middle of the night awakening would be tested in this order: TST based on PSG (Polysomnography), Sleep efficiency based on PSG, Subjective Sleep quality rating, Subjective Sleep Onset Latency, Subjective TST, and Latency to Persistent Sleep based on PSG. The last one in the list is for the low dose vs. placebo comparison and the rest are for the high dose vs. placebo comparison. This reviewer verified that the prespecified analysis for each of these secondary endpoints was significant at the 0.05 level. Note that Sleep Efficiency endpoint is proportional to TST/(Time in Bed). In this trial the Pearson correlation between TST and sleep efficiency is estimated by this reviewer to be 0.999. Therefore, in this reviewer's opinion the Sleep Efficiency endpoint does not seem to provide any distinct information from the TST endpoint.

In study ZI-12 there was a bit of an imbalance at baseline in TST post MOTN awakening, which was the highest secondary endpoint in the hierarchy: 241.6 for Zolpidem and 223.3 for placebo,  $p=0.035$ . However, if we consider that many variables were compared at baseline then after adjusting for multiplicity this would not be statistically significant and the analysis of TST was adjusted for baseline TST (baseline TST was included as a covariate). The statistical test should still be valid despite the possible baseline imbalance and it did not find a statistically significant difference between the treatment groups in terms of TST after MOTN awakening,  $p=0.128$ . Since the TST was the first secondary endpoint in the prespecified testing hierarchy and the test was not statistically significant the results for the subjective WASO (Wake Time after Sleep Onset) and subjective number of awakenings, the other prespecified key secondary endpoints, should be considered exploratory in accordance with the prespecified multiplicity adjustment plan.

There was a slight suggestion in both studies that the treatment effect on Latency increased somewhat with age. However, if we focused on the first period of the crossover study then no such trend was apparent. Similarly, in study ZI-12, if we compared the treatment group difference in those Age < 55 vs. those Age  $\geq$  55, age groups defined by the sponsor, then the difference was not significantly different between these age groups. Therefore, the interaction effect is debatable and in any case the treatment group difference favored Zolpidem on average numerically for all ages studied.

## 2 INTRODUCTION

### 2.1 Overview

Transcept Pharmaceuticals, Inc. is developing a sublingual low-dose zolpidem tartrate sublingual lozenge for the treatment of patients with insomnia characterized by difficulty returning to sleep after awakening in the middle of the night (MOTN). The IND number associated with the development of this drug for the proposed indication is 69,209. The active agent in the sublingual lozenge is zolpidem tartrate, a non-benzodiazepine hypnotic agent of the imidazopyridine class of drugs. Zolpidem tartrate is a Schedule IV controlled substance and is the active ingredient of Ambien® (10 mg for adults <65 years of age, 5 mg for elderly adults)

and Ambien® CR (12.5 mg for adults, 6.25 mg for elderly adults) oral tablet formulations approved and marketed for treatment of insomnia. These zolpidem tartrate formulations require 7–8 hours post-dose time in bed. Ambien® and Ambien® CR, therefore, are not suitable for middle-of-the-night dosing, due to their prolonged duration of action.

The clinical efficacy data on zolpidem tartrate sublingual lozenge are based on studies in 377 patients (118 male, 259 female). The evaluation of efficacy of the zolpidem tartrate sublingual lozenge for insomnia characterized by difficulty returning to sleep after awakening in the middle of the night relies on the findings of two pivotal studies. Study ZI-06-010, a PSG study, employed a scheduled awakening model of middle-of-the-night insomnia to simulate middle-of-the-night awakenings. Study ZI-12, a 4-week double-blind, placebo-controlled, self-report outpatient study, enrolled patients with primary insomnia characterized by spontaneous middle-of-the-night awakenings with difficulty returning to sleep. Note that these studies did not include any geriatric patients but a pharmacokinetic study, ZI-14, was done in the geriatric population.

## 2.2 Data Sources

The network location for the current submission is \CDSESUB1\EVSPROD\NDA022328\0000. The study reports are located in the following directory:

<\\CdseSub1\evsprod\NDA022328\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\insomnia\5351-stud-rep-contr>

The datasets for the key efficacy studies were located as follows at the time of this review:

<\\cdseSub1\EVSPROD\NDA022328\0000\m5\datasets\zi-06-010>  
<//cdseSub1/EVSPROD/NDA022328/0000/m5/datasets/zi-12/>.

## 3 STATISTICAL EVALUATION

### 3.1 Evaluation of Efficacy

#### 3.1.1 Study ZI-06-010

The study was conducted at five United States sleep laboratory sites located in Atlanta (2 labs), Georgia, Cincinnati, Ohio, Overland Park, Kansas, and San Diego, California. The trial took place between 04 April 2006 and 30 July 2006.

##### 3.1.1.1 Study Plan

##### Primary Objective

The primary objective was to evaluate the objective and patient-reported measures of sleep onset of sublingually administered zolpidem versus placebo after MOTN awakening on two

consecutive nights in adult insomnia patients who have difficulty returning to sleep after MOTN awakening.

### Study Design

This was a multi-center, randomized, double blind, placebo-controlled, PSG sleep laboratory, and 3- period crossover study. After initial screening including completion of a 10-day sleep diary, patients were to undergo PSG screening to assess eligibility during which patients were to receive single-blind placebo for 2 consecutive nights after a scheduled MOTN awakening to assess patient eligibility.

Eligible patients were to be randomized to one fixed treatment sequence consisting of 3 periods in accordance with a predetermined randomization schedule, whereby patients were to receive either sublingual zolpidem 1.75 mg, 3.5 mg, or placebo during each treatment period.

Each treatment period was to consist of 2 consecutive study nights in the sleep laboratory, separated by 5 to 12 washout days between each treatment period.

Double-blind study drug was to be administered after a scheduled MOTN awakening for 2 consecutive nights. Patients were to be awakened at 4 hours after initial lights-out, receive study drug, complete a *MOTN Awakening Questionnaire*, and were to be kept awake for 30 minutes before returning to bed to sleep for 4 more hours. Thus, the PSG was to be recorded for a total of 8 hours, though interrupted by a 30-minute break in the middle of the PSG recording.

As soon as possible after the second scheduled awakening and following toilet and dress (30 minutes), patients were to complete a *Treatment Morning Sleep Questionnaire*.

Patients were to be given the *10-day Screening Morning Sleep Diary* and were to be instructed to complete the diary questions each morning for a total of 10 days. At least 7 days of the *10-day Screening Morning Sleep Diary* must have been completed. The diary was to establish eligibility based on subjective MOTN criteria and was also used to ascertain the patient's usual bedtime. To be eligible for randomization patients had to report during the last 7 days recorded in the diary a usual bedtime between 2200 (10:00 PM) and midnight, and awakenings  $\geq 3$  nights per week with a mean SOL (Sleep Onset Latency) of  $\geq 30$  minutes post-awakening.

Patients who continued to be eligible upon completion of the *10-day Screening Morning Sleep Diary* were to undergo PSG screening for MOTN insomnia. This was to consist of 2 consecutive nights with single-blind placebo treatment scheduled as close as possible to randomization. Patients must have had a mean LPS  $\geq 20$  minutes after scheduled MOTN awakenings for both nights and neither night LPS  $< 15$  minutes to remain eligible.

### Study Timepoints

*Visit 1:* Screening and study entry.

*Visits 2 and 3:* Single-blind placebo and PSG screening (baseline) on 2 consecutive nights.

*Visits 4 and 5:* (Treatment Period 1): Patients were randomized to a predetermined treatment sequence and received the first double-blind study drug during 2 consecutive nights.

*Visits 6 and 7:* (Treatment Period 2): Patients received the second double-blind study drug of the treatment sequence to which they were randomized during 2 consecutive nights.

*Visits 8 and 9:* (Treatment Period 3): Patients received the third and final double-blind study drug of the treatment sequence to which they were randomized during 2 consecutive nights. There was a 5- to 12-day washout interval between each of the three treatment periods.

## STUDY POPULATIONS

**Safety Population:** All patients who were randomized into the study and received at least one dose of study drug. This population was to be used for all safety analyses.

**Modified Intent to Treat (mITT) Population:** Efficacy analyses were to be based on a modified intent-to-treat (mITT) population, defined as those patients in the safety population who had LPS data from at least one night of one treatment period. Patients included in the mITT population were to be analyzed in the treatment groups to which they were randomized.

**Latency to Persistent Sleep (LPS)** = number of epochs from the beginning of the recording (following MOTN awakening) to the start of the first 20 non-wake minutes divided by 2.

**Number of Awakenings (NAW) after study drug** = number of periods of awakening of 1 minute or longer (i.e., at least 2 epochs in duration) after onset of persistent sleep following MOTN awakening. Each pair of awakenings must be separated by a Stage 2, 3, 4, or rapid eye movement (REM). In addition to the PSG measurement of NAW there will also be a subjective answer for NAW from the Treatment Morning Sleep Questionnaire.

### Primary Efficacy Variable

- Average LPS after scheduled MOTN awakening (3.5 mg TO (transoral) zolpidem versus placebo)

### Secondary Efficacy Variables

Secondary Polysomnography and patient morning sleep questionnaire variables following MOTN awakening and study drug administration include:

#### Polysomnographic Variables:

- Average LPS after scheduled MOTN awakening (1.75 mg TO zolpidem versus placebo)
- Average Total Sleep Time (TST) after scheduled MOTN awakening (3.5 mg TO zolpidem versus placebo)
- Average Sleep Efficiency (SE) after scheduled MOTN awakening (3.5 mg TO zolpidem versus placebo)

#### Subjective Morning Sleep Questionnaire Variables

- Average Sleep Onset Latency (SOL) after scheduled MOTN awakening (3.5 mg TO zolpidem versus placebo)
- Rating (Poor, Fair, Good, or Excellent) of Sleep Quality after scheduled MOTN awakening (3.5 mg TO zolpidem versus placebo)
- Average Subjective TST (sTST) after scheduled MOTN awakening (3.5 mg TO zolpidem versus placebo)

The secondary efficacy variables and their order of testing are to be as follows:

1. TST after MOTN awakening based on PSG (TO zolpidem 3.5 mg versus placebo)

2. Sleep efficiency after MOTN awakening based on PSG (TO zolpidem 3.5 mg versus placebo)
3. Rating of sleep quality post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
4. SOL post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
5. Subjective TST post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
6. LPS based on PSG after MOTN awakening (TO zolpidem 1.75 mg versus placebo)

If the test of a secondary endpoint is not statistically significant, then inferential analyses of secondary endpoints were to cease and no further inferential assessment of the remaining secondary endpoints were to be made.

All efficacy endpoints not tested as part of the testing hierarchy were to be considered exploratory. A 0.05 significance level was to be used.

### **Determination of Sample Size**

A sample size of 78 was expected to have greater than 90% power to detect a difference in mean LPS following MOTN awakening of 9 minutes (e.g., a TO zolpidem 3.5 mg population mean of 36 minutes and a placebo population mean of 45 minutes), assuming a standard deviation of differences of 24 minutes, using a paired t-test with a 0.05 two-sided significance level. Up to 84 patients were to be enrolled in the study to ensure 78 subjects had LPS following MOTN awakening from at least one night of the TO zolpidem 3.5 mg and the placebo treatment periods.

### **Handling of Dropouts and Missing Data**

Patients who discontinue the study early were not to be replaced. Patients that withdrew early from the study before the final scheduled visit were to have an Early Termination visit. For most endpoints, the data were to be collected on two nights during each treatment period and the mean of the data from the two nights was to be used for analysis. If the data for one night was missing, the values for the other night were to be used. If both nights were missing, then the observation for the treatment period was to be set to missing. For the morning categorical ratings of sleep quality, level of refreshed sleep, and ability to function, the lowest morning rating made during each treatment period was to be used for the categorical ratings.

### **Statistical Analysis Plan**

The primary efficacy variable is the average LPS after scheduled MOTN awakening for TO zolpidem 3.5 mg compared to placebo. The null hypothesis of no differences between treatments will be tested. The primary analysis for testing sleep latency will be the analysis of covariance (ANCOVA) model for a 3-treatment crossover study, including fixed effect terms for Sequence, Period, and Treatment, and a random effect term of patient within sequence for the mITT population. Total Sleep Time during the 4-hour period prior to MOTN awakening (baseline TST) will be used as a covariate in the model, where the total sleep time prior to MOTN awakening for each treatment period is the mean of the 2 total sleep time observations prior to MOTN awakening during the treatment period. The ANCOVA model will assume that TST during the

4-hour period prior to MOTN is not affected by previously administered treatments. Initially, baseline TST by treatment interaction will be included in the model and tested. In the case of non-significance (10% level), it will be dropped from the model, and baseline TST will be assumed to be consistent across treatments (i.e., there is no covariate by treatment interaction).

Let AVE\_LPS denote the average of the LPS over the two nights of a given crossover period. The following SAS code will be used to fit the primary analysis model.

Proc mixed;

Class period treatment sequence subjid;

Model AVE\_LPS= period sequence treatment tst treatment\*tst / ddfm=kr solution;

Random subjid(sequence);

Repeated period / type = CS subject=subjid(sequence);

where tst is the average baseline TST and treatment\*tst is included to test if the covariate effects are consistent across treatments.

The above SAS code will be run 6 different times, with the following covariance structures assumed by substituting the following code in the Repeated statement for “type = CS” on the second through sixth runs:

type = Csh heterogeneous variance compound symmetry

type = Arh(1) heterogeneous first order auto regressive

type = Ante(1) first order ante-dependence

type = TOEP Toeplitz

type = Un unstructured

After fitting the model 6 different times, the model with the covariance structure that has the smallest Akaike Information Criterion (AIC) will be selected.

If the “treatment\*tst” term is not significant in the model chosen (0.10 level of significance), then it will be dropped from the model.

The mean baseline TST based upon PSG from the first 4 hours of sleep during each treatment period also will be computed and will be used as a covariate in analysis. For each endpoint, when the value for one night is missing, the value for the other night will be used. If both nights are missing, then the observation for the treatment period will be set to missing. Since there are three treatment periods each with two consecutive nights of recordings, there are three different mean baseline TST observations, one for each treatment period. As the first step in analysis, a crossover analysis ([Milliken and Johnson, 1992](#)) using an AR(1) covariance structure of the mean baseline TST observations will be used to confirm that there are no treatment mean differences.

ANCOVA will be performed on logarithmic transformed data for LPS and SOL. Natural logarithmic transformation of sleep latency is planned, as the distribution of latency is typically skewed and more variable with placebo than with zolpidem. The same analytical technique (i.e., ANCOVA with TST during the 4-hour period prior to MOTN awakening as the covariate) described for the primary efficacy variable will be used for other PSG endpoints and for continuous morning sleep questionnaire variables. Where necessary to satisfy analysis assumptions, data will be either transformed prior to analysis or a nonparametric procedure will be used.

Two supplemental analyses are planned for the sleep latency endpoints. An additional analysis for testing LPS and SOL will be a standard analysis of variance (ANOVA) model for a 3-treatment crossover study, including fixed effect terms for Sequence, Period, and Treatment, and a random effect term of subject within sequence for the mITT population. The ANOVA also will be performed on logarithmic transformed data. In addition, LPS and TST after study drug administration will be analyzed using analysis of covariance in the subset of patients who had wake time greater than or equal to an average of 60 minutes post MOTN awakening at baseline.

Categorical morning sleep questionnaire results include the following: The rating of sleep quality, rating of level of refreshed sleep and rating of ability to function, the frequency counts will be summarized at baseline and for each treatment within period. The Cochran-Mantel-Haenszel test stratified by period with interval scoring will be performed to test treatment effects for categorical variables.

Evaluation of DSST (Digit Symbol Substitution Test) and VAS (Visual Analog Scale for sedation/alertness) will be based on the mean of the two observations from each treatment period. If the data for one night are missing, the values for the other night will be used. If both nights are missing, then the observation for the treatment period will be set to missing. Mean residual sedation variables will be analyzed in the same manner as the primary efficacy variable. Where necessary to satisfy analysis assumptions, the mean DSST and VAS data may be transformed prior to analysis.

### **3.1.1.2 Demographics and Baseline Characteristics**

A total of 83 patient volunteers were enrolled. Of these, 82 patients (58 female and 24 male) were randomized to treatment. Their median age (range) was 46.7 (19-64) years. Ethnic distribution was 51.2% white, 43.9% black, and the remaining 4.8% was equally divided between Hispanic and Asian.

Table 1 Study ZI-06-010: Demographics and Baseline Characteristics

Variable	Sequence	Sequence	Sequence	Sequence	Sequence	Sequence	
Statistic	ABC	ACB	BAC	BCA	CAB	CBA	Total
<b>Age (yrs)</b>							
N	14	13	13	15	13	14	82
Mean	48.2	47.2	49.6	45.0	41.8	43.5	45.9
(SD)	(10.21)	(14.90)	(10.74)	( 9.64)	(13.33)	(12.99)	(11.97)
Median	51.1	50.3	49.7	43.0	47.3	42.7	46.7
Min, Max	25, 60	19, 64	29, 64	29, 63	19, 57	22, 64	19, 64
<b>Gender</b>							
Male	6 (42.9%)	3 (23.1%)	3 (23.1%)	4 (26.7%)	3 (23.1%)	5 (35.7%)	24 (29.3%)
Female	8 (57.1%)	10 (76.9%)	10 (76.9%)	11 (73.3%)	10 (76.9%)	9 (64.3%)	58 (70.7%)
<b>Number of Nights with MOTN Awakening<sup>a</sup></b>							
N	14	13	13	15	13	14	82
Mean	9.3	9.4	8.9	9.3	9.3	8.6	9.1
(SD)	( 0.99)	( 1.19)	( 1.38)	( 0.98)	( 1.11)	( 1.50)	( 1.20)
Median	10.0	10.0	10.0	10.0	10.0	9.0	10.0
Min, Max	7, 10	7, 10	6, 10	7, 10	7, 10	5, 10	5, 10



Table 1 Study ZI-06-010: Demographics and Baseline Characteristics (continued)

Variable	Sequence	Sequence	Sequence	Sequence	Sequence	Sequence	
Statistic	ABC	ACB	BAC	BCA	CAB	CBA	Total
<b>Race</b>							
Caucasian	8 (57.1%)	8 (61.5%)	7 (53.8%)	6 (40.0%)	7 (53.8%)	6 (42.9%)	42 (51.2%)
Hispanic	1 ( 7.1%)	0 ( 0.0%)	1 ( 7.7%)	0 ( 0.0%)	0 ( 0.0%)	0 ( 0.0%)	2 ( 2.4%)
Black	5 (35.7%)	5 (38.5%)	4 (30.8%)	9 (60.0%)	6 (46.2%)	7 (50.0%)	36 (43.9%)
Asian/ Pacific	0 ( 0.0%)	0 ( 0.0%)	1 ( 7.7%)	0 ( 0.0%)	0 ( 0.0%)	1 ( 7.1%)	2 ( 2.4%)
<b>Weight (kg)</b>							
N	14	13	13	15	13	14	82
Mean	84.41	75.10	81.83	77.46	72.91	77.50	78.25
(SD)	(11.409)	(12.665)	(17.816)	(16.739)	(14.421)	(10.331)	(14.249)
Median	85.60	74.80	78.90	74.70	67.70	78.80	76.85
Min, Max	62.3, 110.0	52.3, 95.9	56.6, 110.9	58.6, 111.4	48.6, 91.5	61.4, 94.5	48.6, 111.4
<b>Height</b>							
N	14	13	13	15	13	14	82
Mean	174.6	166.4	172.4	167.2	167.4	169.8	169.6
(SD)	( 9.353)	( 7.574)	(10.740)	(10.246)	( 8.775)	( 9.955)	( 9.700)
Median	172.9	167.6	172.7	163.8	167.6	168.9	167.8
Min, Max	160.0, 188.0	147.3, 175.3	152.0, 188.0	159.0, 190.5	152.4, 185.0	152.0, 185.4	147.3, 190.5
<b>BMI</b>							
N	14	13	13	15	13	14	82
Mean	27.72	27.10	27.27	27.43	25.92	26.84	27.06
(SD)	( 3.494)	( 4.164)	( 4.164)	( 3.717)	( 4.393)	( 2.681)	( 3.718)
Median	26.55	28.80	26.60	27.40	24.90	27.90	26.75
Min, Max	22.2, 33.4	18.6, 32.1	20.4, 33.3	22.8, 33.5	20.4, 33.6	21.7, 30.0	18.6, 33.6

Note: A = Placebo; B = Zolpidem tartrate lozenge 3.5 mg; C = Zolpidem tartrate lozenge 1.75 mg.

### 3.1.1.3 Patient Disposition

A total of 83 patient volunteers were enrolled and of these, 82 patients were randomized, received at least 1 dose of study drug during the double-blind treatment period, and were included in safety and efficacy analyses. Patient 541 completed through Visit 3 but withdrew consent prior to randomization. Thus, the Safety population and mITT Population were identical. A total of 80 patients completed the study. One patient was discontinued for withdrawal of consent and the other for a family emergency.

**Table 2 Study ZI-06-010: Patient Disposition**

<b>Population</b>	<b>Sequence ABC</b>	<b>Sequence ACB</b>	<b>Sequence BAC</b>	<b>Sequence BCA</b>	<b>Sequence CAB</b>	<b>Sequence CBA</b>	<b>Total</b>
Randomized <sup>a</sup>	14	13	13	15	13	14	82
Safety Population <sup>b</sup>	14(100%)	13(100%)	13(100%)	15(100%)	13(100%)	14(100%)	82(100%)
Modified Intent-to-Treat (mITT) Population <sup>c</sup>	14(100%)	13(100%)	13(100%)	15(100%)	13(100%)	14(100%)	82(100%)
Patients with Baseline Total Awake Time $\geq$ 60 min <sup>d</sup>	10(71.4%)	8(61.5%)	6(46.2%)	2(13.3%)	4(30.8%)	8(57.1%)	38(46.3%)
Completed Study	14(100%)	13(100%)	13(100%)	15(100%)	12(92.3%)	13(92.9%)	80(97.6%)
Discontinued	0( 0.0%)	0( 0.0%)	0( 0.0%)	0( 0.0%)	1( 7.7%)	1( 7.1%)	2( 2.4%)
Reason for Discontinuation							
Other	0( 0.0%)	0( 0.0%)	0( 0.0%)	0( 0.0%)	0( 0.0%)	1( 7.1%)	1(1.2%)
Withdrawal Of Consent	0( 0.0%)	0( 0.0%)	0( 0.0%)	0( 0.0%)	1( 7.7%)	0( 0.0%)	1(1.2%)

<sup>a</sup> All patients assigned a randomization number. Patient 541 withdrew consent prior to randomization. <sup>b</sup> All randomized patients who received at least 1 dose of study medication. <sup>c</sup> All Safety Population patients who had LPS data from at least 1 night from at least 1 treatment period. <sup>d</sup> Patients with Baseline Awake Time  $\geq$  60 minutes are all mITT patients with average Screening Period (baseline) post-MOTN Awake Time  $\geq$  60 minutes. Note: Percentages are based on the number randomized. Note: A = Placebo; B = zolpidem tartrate lozenge 3.5 mg; C = zolpidem tartrate lozenge 1.75 mg.

Patients 104 and 503 completed the first visit of the treatment period but were not able to return for the second night. Both patients later restarted the treatment period to ensure dosing over two consecutive days and as a result received an additional dose of study drug.

### 3.1.1.4 Sponsor's Results

Dosing errors were recorded for seven patients. None of these errors resulted in the exclusion of data from analyses.

**Table 3 Study ZI-06-010 Dosing Errors**

SUBJECT	PERIOD	VISIT(S)	COMMENT
215	3	6, 7	Dosing Error Randomized to receive Treatment A (placebo) and received Treatment C (zolpidem tartrate lozenge 1.75 mg) both nights.
211	3	8, 9	Dosing Error Randomized to receive Treatment B (zolpidem tartrate lozenge 3.5 mg) and received Treatment C(zolpidem tartrate lozenge 1.75 mg) at Visit 8.
515	3	9	Dosing Error Treatment B (zolpidem tartrate lozenge 3.5 mg) given instead of scheduled Treatment C (zolpidemtartrate lozenge 1.75 mg).

Compared to placebo, zolpidem tartrate lozenges produced a significant dose-related decrease in LPSMOTN (LPS post MOTN awakening). Adjusted LS means for LPSMOTN in the mITT population were 28.1 minutes after placebo, 16.9 minutes after zolpidem tartrate lozenge 1.75 mg ( $P<0.001$ ), and 9.7 minutes after zolpidem tartrate lozenge 3.5 mg ( $P<0.001$ ) as presented in Table 4. Results of these analyses demonstrated that difference in LPSMOTN between doses was also statistically significant ( $P<0.001$  [note: exploratory analysis unadjusted for multiplicity]).

**Table 4 Study ZI-06-010: Sponsor's Results for LPS Post MOTN Awakening**

PSG Variable LPSMOTN (min)	Screening	3.5 mg	1.75 mg	Placebo	P-value
<b>Day 1, Post MOTN Awakening</b>					
N	82	79	82	81	
Mean (SEM)	54.66 (3.718)	14.31 (1.899)	24.59(3.292)	36.88(3.714)	
Median	44.75	8.50	17.50	30.00	
Min, Max	15.0, 177.5	0.0, 87.0	0.5, 238.5	2.5, 240.0	
<b>Day 2, Post MOTN Awakening</b>					
N	80	80	82	81	
Mean (SEM)	44.18 (2.882)	13.91 (1.469)	22.84(3.327)	37.70(3.690)	
Median	40.00	10.25	14.25	26.50	
Min, Max	10.5, 120.5	0.0, 70.0	0.5, 240.0	3.5, 158.5	
<b>Mean Days 1 and 2, Post MOTN Awakening</b>					
N	82	80	82	81	
Mean (SEM)	49.83 (2.879)	14.16 (1.378)	23.71(2.650)	37.29(3.016)	
Median	45.00	11.25	16.75	29.25	
Min, Max	19.8, 149.0	1.0, 60.5	2.3, 145.5	3.3, 136.0	
<b>ANCOVA Analysis<sub>a</sub></b>					
LS Mean		9.69	16.89	28.12	Period: 0.711
95% CI		8.06, 11.65	14.07, 20.26	23.41, 33.77	Sequence: 0.984
					Treatment: <0.001
<b>Comparison to Placebo<sub>b</sub></b>					
Difference in LS Means		0.34	0.60		
95% CI for Difference		0.28, 0.42	0.49, 0.74		
P-value		<0.001	<0.001		
<b>Comparison of zolpidem tartrate lozenge 3.5 mg to 1.75 mg<sub>b</sub></b>					
Difference in LS Means		0.57			
95% CI for Difference		0.47, 0.70			
P-value		<0.001			

### 3.1.1.5 Reviewer's Results

#### 3.1.1.5.1 Primary Analysis

Based on the primary analysis the 3.5 mg group had a significantly lower mean log LPS than placebo. The estimated difference was  $1.06 \pm 0.104$  (on the log scale),  $p < 0.0001$ . The difference between 1.75 mg and placebo was estimated to be  $0.509 \pm 0.103$ ,  $p < 0.0001$ . Differences between 3.5 mg and placebo in LPS were roughly consistent from period to period. Differences between 1.75 mg and placebo were slightly less consistent. In particular, the estimated difference between 1.75 mg and placebo in period 2 was numerically smaller than that in period 1 or period 3. However, there was no compelling evidence of a treatment by period interaction or carryover effect. The period effect was also not statistically significant ( $p = 0.710$ ). For example, the estimated mean log LPS was 3.40, 3.17, and  $3.43 \pm 0.15$  for those on placebo in periods 1, 2, and 3, respectively.

#### 3.1.1.5.2 Analysis of Crossover Period 1 Only

Tests based on only data from the first period supported those based on all three periods in terms of whether or not statistical significance was achieved (3.5mg vs Pl difference:  $-1.05 \pm 0.23$   $p < 0.0001$ ; 1.75 mg vs Pl:  $-0.50 \pm 0.23$   $p = 0.0321$ ). Transformed back from the log scale to the original LPS scale the group mean estimates are 28.12, 16.89, and 9.69 minutes for placebo, 1.75 mg, and 3.5 mg, respectively.

#### 3.1.1.5.3 Investigation of the Impact of Missing Data on the Primary Analysis Result

In the following notation C represents 1.75 mg, B represents 3.5 mg, and A represents placebo. Two patients completed just one or two of the three treatment periods. One such patient was randomized to the CAB sequence and only completed C and A; the other was randomized to CBA and only completed C. The worst LPS was 136 for the placebo group, 145.5 for group C (1.75mg) and 60.5 for group B (3.5mg). Therefore, although it would not be the worst possible scenario, imputing a LPS of 136 for period 3 (1.75 mg) for the first patient and 136 for period 2 (1.75 mg) and 0 for period 3 (placebo) for the second patient would represent an unfavorable imputation similar to a worst case scenario for the drug groups. If we impute missing LPS scores (post MOTN awakening) for the two individuals who did not complete all three periods in this manner then both groups are still statistically significantly better than placebo in terms of the log LPS after middle of the night awakening ( $p = 0.0001$  for 1.75 vs. placebo and  $p < 0.0001$  for 3.5 vs. placebo). After transforming the estimates of log LPS back to the original non-logarithm LPS scale the group mean estimates are 26.9, 16.9, and 10.3 minutes based on the data incorporating this worst case imputation. Based on this it seems that

the missing data were it available would not likely alter the significance of the primary analysis result.

#### ***3.1.1.5.4 Comparison of LPS pre MOTN awakening across Groups and Crossover Periods***

It seems that the LPS for the first four hours before the scheduled awakening is less consistent than one would expect across the treatment groups as well as across the periods. Because it covers the time before administration of treatment there should be no difference between the treatment groups and the same goes for the crossover period if there is no carryover effect from one period to the next. However, this reviewer's analysis based on a similar model to the primary analysis model (only excluding the baseline as a covariate because it is the response variable in this case) suggests that 3.5 mg LPS is higher on average than placebo and also that the LPS before scheduled awakening in crossover period 3 is on average lower than that in crossover period 1 or 2. Table 5 presents the mean LPS by crossover period and assigned treatment in that period. The period 3 mean for those assigned placebo in period 3 is notably smaller than for those assigned placebo in period 1 or 2. From

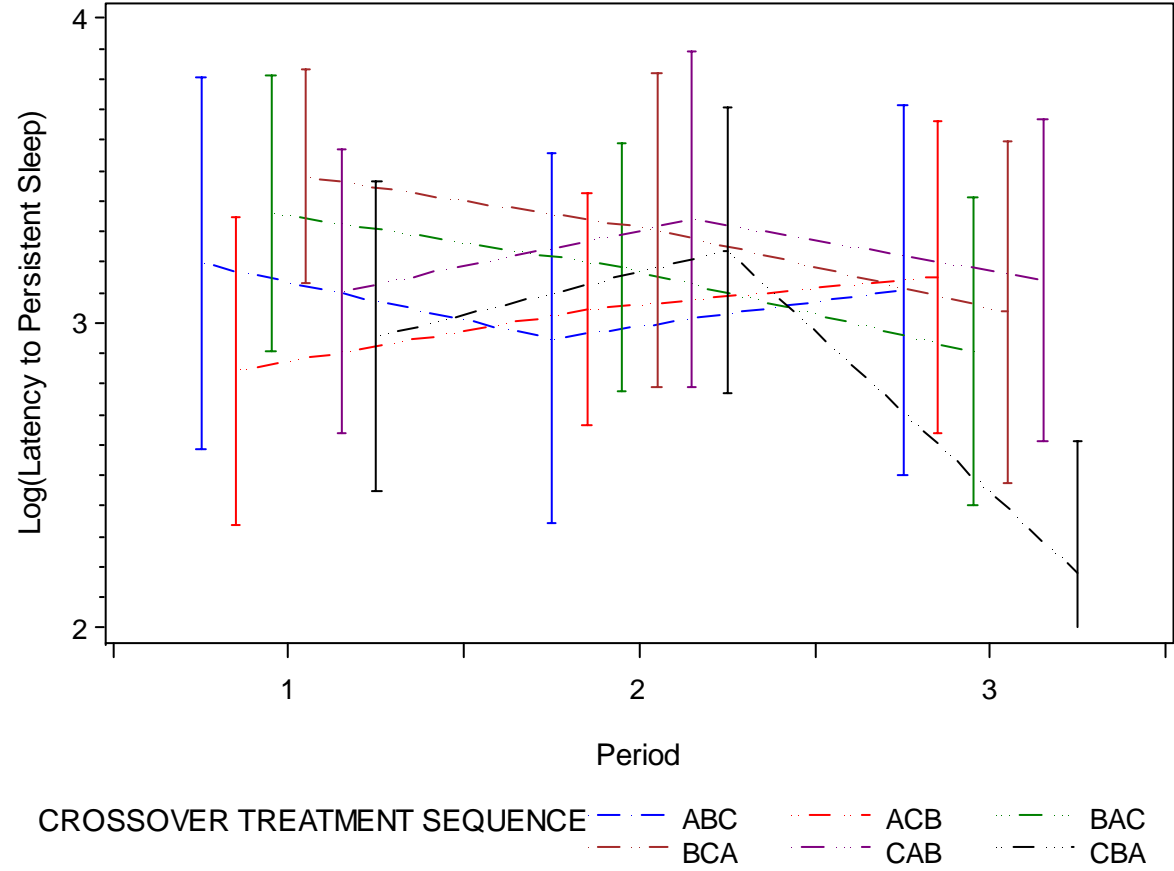
Appears This Way On  
Original

Figure 1 we can see that the significance may be driven by the patients in the sequence CBA (1.75/3.5/Placebo) whose mean in period 3 seems to be much lower than the others. Note that in the figure the groups have been artificially spread out along the x- axis to avoid having the mean +/- 2 standard error bars being on top of each other.

Table 5 Study ZI-06-010: LPS before Scheduled Awakening

	D.B. Med					
	Placebo		3.5 mg		1.75 mg	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
<b>CROSSOVER PERIOD</b>	21.32	3.49	28.69	4.62	21.43	3.51
<b>1</b>						
<b>2</b>	26.63	4.43	23.31	3.81	22.75	3.67
<b>3</b>	13.54	2.18	23.56	3.99	20.81	3.41

Figure 1 Study ZI-06-010: LPS before MOTN awakening (before drug administration)



To avoid this problem we may consider focusing on period 1 data. Just looking at period 1 may not solve the problem because the LPS before scheduled awakening is numerically and nearly nominally significantly higher on average for those who got 3.5 mg in period 1 than for those who got placebo or 1.75 mg. This is all based on the average over the 2 days in each crossover period as prespecified. The difference in LPS before scheduled awakening seems to be driven mostly by the 2<sup>nd</sup> day of each crossover period. If we just analyze the first day of each crossover period then the differences in LPS before scheduled awakening between treatment groups become insignificant when averaged over all crossover periods. However, if we focus on period 1 alone there still appears to be a difference with the 3.5 group being higher on average than the other 2 groups (3.5 vs. 1.75:  $p=0.0375$ ; 3.5 vs. placebo/1.75 combined:  $p=0.0473$ ). If there are in fact significant differences at baseline then the differences seen after treatment may not be attributable to the treatment so that the estimate of treatment difference may be biased.

Another approach to deal with this problem uses the period 0 (single blind placebo) LPS as the baseline for each crossover period. There were no differences between the sequences in mean period 0 LPS before MOTN awakening. The analysis of LPS post MOTN awakening adjusted for period 0 LPS before MOTN awakening suggests that LPS post MOTN awakening was significantly lower on average for 3.5 mg than placebo ( $p<0.0001$ ) as well as for 1.75 mg compared to placebo ( $p<0.0001$ ). The estimated differences from placebo are 18.6 and 9.6 minutes for 3.5 mg and 1.75 mg, respectively.

The LPS before MOTN awakening on night 2 of each crossover period would have some possibility of being affected by the previous night's treatment. Table 6 shows the mean LPS before scheduled awakening on Day 1 of each crossover period and Table 7 shows it for Day 2. If we focus on LPS before MOTN awakening on night 1 of each crossover period then globally there are no significant period or drug group assignment effects (globally meaning simultaneously testing for any differences among the three drug groups rather than looking at all pairwise comparisons).

Table 6 Study ZI-06-010: LPS before Scheduled Awakening Day 1 of Crossover Period

	<b>D.B. Med</b>					
	<b>Placebo</b>		<b>3.5 mg</b>		<b>1.75 mg</b>	
	<b>Mean</b>	<b>S.E.</b>	<b>Mean</b>	<b>S.E.</b>	<b>Mean</b>	<b>S.E.</b>
<b>CROSSOVER PERIOD</b>						
<b>1</b>	19.33	3.17	27.39	4.42	18.20	3.04
<b>2</b>	21.91	3.65	21.35	3.50	26.68	4.31
<b>3</b>	21.57	3.48	19.51	3.35	16.75	2.74



Table 7 Study ZI-06-010: LPS before Scheduled Awakening Day 2 of Crossover Period

	<b>D.B. Med</b>					
	<b>Placebo</b>		<b>3.5 mg</b>		<b>1.75 mg</b>	
	<b>Mean</b>	<b>S.E.</b>	<b>Mean</b>	<b>S.E.</b>	<b>Mean</b>	<b>S.E.</b>
<b>CROSSOVER PERIOD</b>						
<b>1</b>	21.32	3.49	28.69	4.62	21.43	3.51
<b>2</b>	26.63	4.43	23.31	3.81	22.75	3.67
<b>3</b>	13.54	2.18	23.56	3.99	20.81	3.41

The analysis of LPS after scheduled awakening based on the first day of each crossover period is consistent with that based on the average of the two days from each crossover period. Therefore, it seems that problem associated with the difference in LPS before scheduled awakening found for the average over the two nights in each crossover period can be circumvented if we are willing to look at just the first day of each crossover period.

In spite of this possible group difference in LPS prior to scheduled awakening the pre-specified primary analysis as well as most other analyses (e.g., restricted to the first of two days in each crossover period or restricted to a particular crossover period) seem to support a significant treatment effect on LPS after scheduled awakening. Therefore, in light of the balance of the data suggesting a robustness of the statistical significance of the treatment effect we may be inclined to disregard the observed group differences in pre-MOTN LPS. Note also that the primary analysis of LPS post MOTN prespecified adjusting for total sleep time before scheduled awakening rather than LPS before scheduled awakening and there were no significant group imbalances in total sleep time before scheduled awakening.

### **3.1.1.5.5 Analyses of Secondary Endpoints**

Total Sleep Time post MOTN as measured by PSG was significantly higher for the 3.5 mg group than placebo. The difference was estimated to be 23.2 minutes ( $p < 0.0001$ ). The 1.75 mg group was also nominally significantly higher than placebo in terms of PSG TST. The estimated difference was 13.8,  $p < 0.0001$ . There was not a significant treatment effect on WASO as measured by PSG (3.5 mg vs. placebo,  $p = 0.6984$ ). The average WASO post MOTN for 3.5 mg group was an estimated 0.846 minutes less than placebo. If we restrict to period 1 data there was still no significant difference between 3.5 mg and placebo in terms of PSG WASO ( $p = 0.9915$ ). Also, over all periods the 1.75 mg was estimated 1.04 minutes higher than placebo,  $p = 0.6324$ .

Mean differences from placebo in subjective TST post MOTN awakening were 22.3 ( $p < 0.0001$ ) for 3.5 mg and 13.2 ( $p = 0.0159$ ) for 1.75 mg. Mean subjective Latency to Sleep Onsets post MOTN awakening were 37.5, 25.3, 27.6, for placebo, 3.5 mg, and 1.75 mg, respectively, ( $p < 0.0001$  for both comparisons with placebo).

The analysis plan stated that secondary endpoints would be tested in the following hierarchical order:

1. TST after MOTN awakening based on PSG (TO zolpidem 3.5 mg versus placebo)
2. Sleep efficiency after MOTN awakening based on PSG (TO zolpidem 3.5 mg versus placebo)
3. Rating of sleep quality post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
4. SOL post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
5. Subjective TST post-dose from the *Treatment Morning Sleep Questionnaire* (TO zolpidem 3.5 mg versus placebo)
6. LPS based on PSG after MOTN awakening (TO zolpidem 1.75 mg versus placebo).

This reviewer verified that the prespecified analysis for each of these secondary endpoints was significant at the 0.05 level.

Note that Sleep Efficiency endpoint is proportional to TST/(Time in Bed). In this trial the Pearson correlation between TST and sleep efficiency is estimated by this reviewer to be 0.999. Therefore, in this reviewer's opinion the Sleep Efficiency endpoint does not seem to provide any distinct information from the TST endpoint.

Subjective WASO was not in the list of key secondary endpoints and was investigated post-hoc by the sponsor.

### **Treatment Effects within Individual Sites**

There were 5 sites that randomized patients in study ZI-06-010. The number of patients per site ranged from 13 to 20 and the number of observations per site ranged from 37 to 60 (1 to 3 per

patient). Analysis of log LPS post MOTN awakening specific to site reached nominal significance for the 3.5 mg vs. placebo comparison for all sites. Treatment differences on the log scale ranged from 0.79 to 1.24. Transformed back to the original scale, in minutes, estimated differences in LS Means ranged from 14.8 to 25.5. For 4 of 5 sites analysis of log LPS post MOTN awakening specific to site reached nominal significance for the 1.75 mg vs. placebo comparison. Transformed back to the original scale, in minutes, estimated differences between LS Means 1.75 mg and placebo ranged from 6.9 to 16.6. Excluding any one particular site did not alter the significance of the result for LPS post MOTN awakening.

### **3.1.2 Study ZI-12**

Study ZI-12 was a Randomized, Double-blind, Placebo-controlled, Parallel Group Study of the Efficacy and Safety of the Zolpidem Tartrate Lozenge in Adult Subjects with Insomnia Characterized by Difficulty Returning to Sleep After Awakening in the Middle-Of-The-Night (MOTN).

The date the first subject enrolled was 16 May 2007 and the date the last subject completed was 26 November 2007. Twenty five sites within the United States participated in the study (30 planned were but 5 did not randomize any subjects).

#### **3.1.2.1 Overall Study Design and Plan**

##### **Primary objective:**

To evaluate the efficacy of the 3.5 mg zolpidem tartrate lozenge versus placebo in reducing latency to sleep onset after spontaneous middle of the night awakenings (LSOMOTN) in adult subjects with sleep maintenance insomnia.

##### **Study Plan**

This was a multicenter, randomized, double-blind, placebo-controlled, parallel group outpatient study that utilized middle-of-the-night dosing with randomized study medication (3.5 mg zolpidem tartrate sublingual lozenge or placebo lozenge) taken on an as needed (prn) basis over 28 nights. During the 2 week screening period, all subjects (n=703) received placebo under single-blind conditions. The subsequent 4-week long randomized treatment comparison phase of the study was conducted under double-blind conditions. Approximately 320 adult male and female subjects were planned to be randomized in order to allow 240 subjects to complete. Eligible subjects were to be randomized on a 1:1 basis to receive either a 3.5 mg zolpidem tartrate lozenge or placebo using a predetermined randomization schedule. Subjects specifically were to be instructed NOT to take the medication at bedtime and were to be instructed to call the IVRS (interactive voice response system) when they had difficulty returning to sleep following a MOTN awakening of at least 10 minutes in duration, provided that the subject would be able to remain in bed for 2 to 4 hours after dosing. During the MOTN IVRS call subjects were to respond to questions concerning their MOTN awakening. After calling the IVRS, if appropriate, subjects were to take the study medication immediately and attempt to go back to sleep.

## **Analysis Plan**

The analysis plan was dated September 8, 2006.

### **SUBJECT POPULATIONS**

The full analysis set of subjects includes all subjects who took at least one-dose of single blind study medication during the placebo run-in period. Safety analyses were to include subjects in the full analysis set who took at least one dose of study drug post-randomization. Subjects included in safety analyses were to be analyzed under the treatment actually received during the double-blind treatment phase of the study. Subjects who were in the full analysis set but excluded from safety analyses were to have their safety information listed but not summarized. Efficacy analyses were to include all subjects who took at least one dose of study drug during the double-blind treatment phase of the study and who had LSOMOTN data from at least one night of the double-blind treatment period. Subjects were to be analyzed in accordance with their assigned treatment based on the randomization schedule.

### **Sample Size Determination**

The average LSOMOTN following dosing for each subject was to be log transformed (natural log) prior to analysis. Consequently, sample size calculations were based upon the expected ratio of 3.5 mg zolpidem tartrate lozenge to placebo. A sample size of 120 subjects in each group was expected to have 90% power to detect a fold change in means (i.e., the expected ratio of 3.5 mg zolpidem tartrate lozenge to placebo) of 0.82 assuming that the coefficient of variation is 0.5 using a two group t-test with a 0.05 two-sided significance level. A sample size of up to 160 subjects per treatment group was selected, as 25% of subjects were expected to drop out of the study prior to the end of Week 4 of treatment period, and this sample size was expected to ensure a total of at least 240 evaluable completers. An evaluable completer was defined as a subject completing all 4 weeks of treatment period and having evaluable IVRS data required for the analyses of primary and secondary endpoints.

### **Pooling of Sites for Statistical Analysis**

Approximately 30 sites in the United States were expected to participate in the study and to enroll a total of 320 subjects. Study sites were to be pooled together regionally to form 15- 20 larger, consolidated sites before performing statistical analyses. Pooling of study sites for analysis was to be done following completion of enrollment in the study but prior to breaking the study blind as follows.

- Sites with 16 or more subjects included in the primary efficacy analysis were not to be pooled for analysis. These were to be stand alone sites.
- All sites with < 16 subjects would be pooled regionally. Within region sites would be ordered from the largest to smallest.
- Sites within each region would be pooled from largest to smallest until each pooled site reaches at least 16 subjects. Pooling would continue until all sites within each region were pooled.
- If the last site within a region had fewer than 16 subjects, then this site would be combined with the immediately prior pooled site.

## **ANALYSIS OF EFFICACY DATA**

All efficacy measures refer to evaluations obtained subsequent to MOTN awakening and self-administration of sublingual study drug. For each measure other than the Insomnia Severity Index (ISI), the mean value for nights on which the subject took study medication during the 4-week treatment period was to be computed. For ISI, the total score at the post treatment Week 4/End of double-blind treatment (range: 0–28) was to be computed.

The baseline value for all efficacy measures other than ISI was the mean for nights during which the subject took study medication during the two-week screening run-in period. For ISI, it was the score for the ISI at the end of the screening run-in period. Differences between treatment groups were to be analyzed using an analysis of covariance model that includes fixed effects for treatment and pooled study site and the baseline value of the outcome variables as a covariate. The analysis was to be performed using all available observations from the 4-week, double blind treatment period, by week and over treatment weeks 1-2 (Day 1 to 14 of treatment period) and 3-4 (Day 15-28) of treatment period. Least squares means (LS means) and 95% confidence intervals for the LS means were to be provided for each of the treatments.

The primary efficacy endpoint is Latency to Sleep Onset post-MOTN awakening (LSOMOTN). Each time a subject takes a dose of study medication, LSOMOTN for that day will be the number of minutes recorded in response to the following IVRS question: "How long did it take you to fall asleep after taking your study medication?" LSOMOTN will be the average of the available LSOMOTN values from nights in which subjects took study medication during the 4-week treatment period. The baseline LSOMOTN will be the average of the available LSOMOTN values from nights in which subjects took study medication during the 2-week placebo single-blind run-in period. The denominator will be the total number of nights with a MOTN awakening and dosing for the time point of interest.

LSOMOTN will undergo a logarithmic transformation prior to analysis. Natural logarithmic transformation of sleep latency is planned, as the distribution of latency is typically skewed and more variable with placebo than with zolpidem.

## **Secondary Efficacy Endpoints**

The following secondary efficacy endpoints were to be analyzed in the same fashion that was performed for LSOMOTN except that no logarithm transformation was to be imposed. A treatment effect was to be evaluated according to the following hierarchical multiple comparison procedure. If the test of treatment group difference is not significantly different, i.e., if p-value is greater than 0.05 for a secondary endpoint, then inferential analyses of secondary endpoints were to cease and assessment of the secondary endpoints remaining in the hierarchy were to be considered exploratory.

The hierarchical order of secondary endpoints was as follows:

1. Subjective Total Sleep Time post MOTN dosing (sTSTMOTN) averaged over all dosing nights during the 4-week Treatment period
2. Subjective Number of Awakenings post MOTN dosing (sNAWMOTN) averaged over all dosing nights during the 4-week Treatment period

3. Subjective Wake Time after Sleep Onset post MOTN dosing (sWASOMOTN) averaged over all dosing nights during the 4-week Treatment period

### **Handling of Missing Data**

In cases where a subject had a qualified MOTN awakening but sTSTMOTN = 0 *or* LSOMOTN = 999 (the subject was instructed to enter 999 if he/she never fell back asleep after taking his/her study medication) , the subject will be assumed to have never fallen back to sleep after MOTN dosing. The following values will be used in analysis:

sTSTMOTN = 0,

LSOMOTN = LSOMOTN, provided LSOMOTN does not equal 999, otherwise,

if LSO equals 999, then LSOMOTN = 4 hours

sNAWMOTN = missing, and

sWASOMOTN = missing

All other data will be analyzed as recorded; no other imputation of data will be carried out. If a subject has no data for a given week (e.g., the subject withdrew from the study prior to the week, did not take study medication and/or failed to call the IVRS), then the results for the week will be missing. In the event that only partial data are available for a given week (i.e., a subject withdraws from the study during the week or fails to call the IVRS), all available observations will be used in analysis for the week.

### **3.1.2.2 Demographic and Other Baseline Characteristics**

Of the 295 subjects in the safety population, 100% had a diagnosis of primary insomnia by the DSM-IV. The median age was 43 years, 68.1% were female, 31.9% were male, 64.4% were white, and 31.2% were Black or African-American. The median weight was 74.4 kg, median height was 169.2 cm, and the median BMI was 26.22. No significant differences in demographics and other baseline characteristics were noted between zolpidem tartrate sublingual lozenge and placebo treatment groups.

**Table 8 Study ZI-12: Baseline Demographics**

Demographic Variable/	3.5 mg	Placebo	Total	P-value
<b>Age (yrs) (n)</b>	150	145	295	
Mean (SD)	42.3 (11.38)	43.4 (11.30)	42.8 (11.33)	0.350 <sup>a</sup>
Median (min, max)	42.0 (18, 64)	44.0 (18, 64)	43.0 (18, 64)	
<b>Gender (n)</b>	150	145	295	
Male	43 (28.7%)	51 (35.2%)	94 (31.9%)	0.203 <sup>b</sup>
Female	107 (71.3%)	94 (64.8%)	201 (68.1%)	
<b>Race (n)</b>	150	145	295	
White	96 (64.0%)	94 (64.8%)	190 (64.4%)	0.919 <sup>b</sup>
Black or African-American	47 (31.3%)	45 (31.0%)	92 (31.2%)	
Asian	3 (2.0%)	2 (1.4%)	5 (1.7%)	
Native Hawaiian or other	1 (0.7%)	0 (0.0%)	1 (0.3%)	
American Indian or Alaska	1 (0.7%)	2 (1.4%)	3 (1.0%)	
Other <sup>c</sup>	2 (1.3%)	2 (1.4%)	4 (1.4%)	
<b>Ethnicity (n)</b>	150	145	295	
Hispanic or Latino	22 (14.7%)	26 (17.9%)	48 (16.3%)	0.420 <sup>b</sup>
Not Hispanic or Latino	128 (85.3%)	119 (82.1%)	247 (83.7%)	
<b>Weight (kg) (n)</b>	150	145	295	
Mean (SD)	74.27 (15.050)	75.73 (13.434)	74.99 (14.273)	0.356 <sup>a</sup>
Median	73.90	74.80	74.40	
<b>BMI (n)</b>	150	145	295	
Mean (SD)	26.17 (3.860)	26.55 (3.754)	26.35 (3.807)	0.428 <sup>a</sup>
Median	26.18	26.25	26.22	

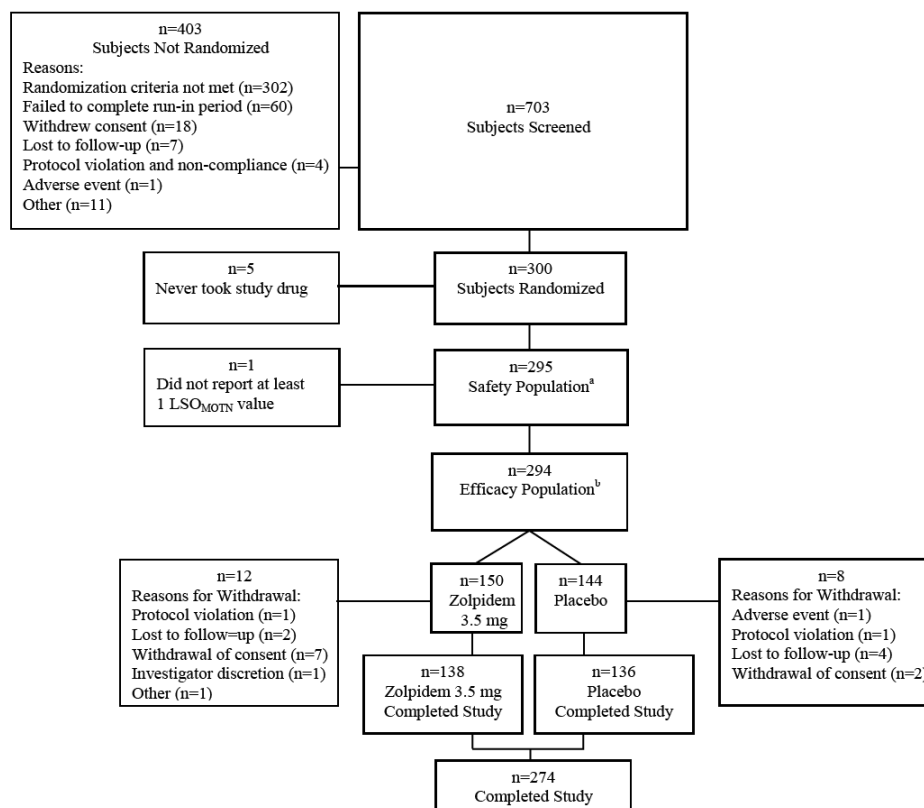
<sup>a</sup>The 2-way ANOVA with factors for treatment and pooled site was used to compare differences between the 2 treatment groups.

<sup>b</sup>The Cochran-Mantel-Haenszel test, stratified by pooled site, was used to test for differences between treatment groups using the general association test.

Copied from page 52 of sponsor's report

### 3.1.2.3 Patient Disposition

Three hundred subjects were randomized and 295 subjects were included in the safety population. Five subjects never took study medication (2 due to protocol violations and 3 due to other reasons).

**Figure 2 Study ZI-12: Disposition of Subjects**

<sup>a</sup>The safety population includes all randomized subjects who took at least 1 dose of study medication during the double-blind treatment phase of the study.

<sup>b</sup>The efficacy population includes all randomized subjects who took at least 1 dose of study medication and had at least 1 LSOMOTN value.

### 3.1.2.4 Sponsor's Results

#### Primary Efficacy Results

The primary efficacy end point was LSOMOTN, defined as the mean of LSO values from nights with post-middle-of-the-night dosing during the 4-week treatment period. Nightly LSOMOTN was the response to the IVRS diary question, "How long did it take you to fall asleep after taking your study medication?" LSOMOTN was compared across treatment groups using an ANCOVA model for logarithmically-transformed LSOMOTN, with the logarithm of baseline mean LSOMOTN, pooled study center, and treatment group as covariates.



Table 9 Study ZI-12: Latency to Sleep Onset Post Middle-of-the-Night Awakening after Dosing

	<b>Zolpidem 3.5 mg</b>	<b>Placebo</b>	<b>P-value</b>
<b>Baseline* (n)</b>	150	144	
Mean (SE)	73.2 (2.82)	75.2 (3.12)	
Median (min, max)	63.8 (27, 234)	64.0 (27, 255)	
LS Mean	68.13	69.42	0.6806a
95% CI of LS Mean	63.90, 72.64	65.03, 74.10	0.2581b
<b>Treatment Week 1 (n)</b>	150	143	
Mean (SE)	50.8 (3.25)	69.4 (3.10)	
Median (min, max)	41.6 (9, 283)	60.8 (13, 185)	
LS Mean	41.03	59.34	<0.0001c
95% CI of LS Mean	37.35, 45.07	53.90, 65.33	0.9744d
<b>Treatment Week 2 (n)</b>	140	139	
Mean (SE)	47.1 (2.89)	66.2 (3.16)	
Median (min, max)	39.3 (9, 201)	60.0 (10, 240)	
LS Mean	37.72	56.04	<0.0001c
95% CI of LS Mean	34.06, 41.78	50.61, 62.05	
<b>Treatment Week 3 (n)</b>	137	134	
Mean (SE)	41.4 (2.48)	62.6 (3.39)	
Median (min, max)	33.3 (4, 180)	52.4 (9, 210)	
LS Mean	33.66	51.14	<0.0001c
95% CI of LS Mean	30.24, 37.47	45.90, 56.98	
<b>Treatment Week 4 (n)</b>	129	130	
Mean (SE)	41.6 (2.67)	62.2 (3.09)	
Median (min, max)	30.8 (5, 180)	55.4 (10, 167)	
LS Mean	32.65	50.93	<0.0001c
95% CI of LS Mean	29.39, 36.27	45.87, 56.55	
<b>DB Treatment (n)</b>	150	144	
Mean (SE)	46.3 (2.61)	64.9 (2.62)	
Median (min, max)	36.8 (10, 206)	59.0 (11, 153)	
LS Mean	38.22	56.37	<0.0001c
95% CI of LS Mean	35.06, 41.66	51.63, 61.56	

\*Baseline LSOMOTN = Average LSOMOTN collected during 2-week placebo single-blind run-in period for nights study medication was taken.

a 2-way ANOVA with log transformed mean LSO as response; treatment and pooled site as fixed effects. P-value is for treatment

b P-value for pooled site-by-treatment interaction when interaction added to model given in a.

c ANCOVA model with log-transformed mean LSO as response; treatment and pooled site as fixed effects and baseline log-transformed average LSO as covariate. P-value shown is for treatment.

d P-value for pooled site-by-treatment interaction when interaction term added to model given in c.

Note 2: LS Means have been backtransformed. The table was copied from the sponsor's study report pages 53-55

The primary end point, LSOMOTN averaged across the 4-week treatment period, was significantly decreased in the zolpidem tartrate sublingual lozenge group versus the placebo group ( $p < 0.0001$ ). The LS Mean LSOMOTN for the zolpidem tartrate sublingual lozenge group at baseline was 68.13 minutes which decreased to 38.22 minutes following the 4-week treatment period. The corresponding LSOMOTN values of the placebo group were 69.42 minutes and 56.37 minutes, respectively.

The difference was significantly in favor of the zolpidem tartrate sublingual lozenge at each of the 4 weeks of the study. After randomization and treatment with study medication, statistically significant improvement (reduction) was seen at Week 1 and the effect persisted throughout the double-blind treatment period for the group treated with zolpidem tartrate sublingual lozenge compared with the placebo group. For Week 1, the LS mean time to return to sleep was 41.03 minutes for subjects that took zolpidem tartrate sublingual lozenge versus 59.34 minutes for subjects that took placebo ( $p < 0.0001$ ). Additionally, nominally significant differences ( $p < 0.0001$ ) were observed between the two treatment groups at Week 2, Week 3, Week 4, Weeks 1 and 2 combined, and Weeks 3 and 4 combined.

## **Secondary Efficacy Results**

### **sTSTMOTN Post-middle-of-the-night Awakening**

Secondary efficacy end points were analyzed in a hierarchical order. The first in the hierarchy of key secondary end points, sTSTMOTN averaged over the 4-week treatment period, was defined as the number of minutes recorded in response to the following IVRS question "After you fell back to sleep, how long did you sleep until you woke up this morning?". sTSTMOTN was directionally positive versus placebo but the difference was not statistically significant ( $p = 0.128$ ). TSTMOTN at baseline for the zolpidem tartrate sublingual lozenge group was 241.2 min (LS Mean) and for the placebo group was 222.9 min (LS Mean). This imbalance between treatment groups at baseline was statistically significant ( $p = 0.0341$ ). Consequently, the correlation of treatment values with baseline was higher for the placebo group (75–83%) than for the zolpidem group (56–61%). This resulted in a lack of statistical significance for this secondary end point. Despite the imbalance between the treatment groups, nominally significant differences in total sTSTMOTN between placebo and zolpidem tartrate sublingual lozenge at Week 1 and Week 2 ( $p = 0.0107$  and  $p = 0.0469$  respectively) in favor of zolpidem tartrate sublingual lozenge were observed, as presented in Table 10. The approximate 23-minute improvement from baseline in the zolpidem tartrate sublingual lozenge group was maintained throughout the 4-week study period.

**Table 10 Study ZI-12: Sponsor's Analysis of Subjective Total Sleep Time**

	<b>Zolpidem 3.5 mg</b>	<b>Placebo</b>	<b>P-value</b>
<b>Number of Subjects</b>	150	144	
<b>Baseline* (n)</b>	150	144	
Mean (SE)	242.0 (5.75)	224.0 (6.57)	
Median (min, max)	240.0 (33, 477)	222.5 (12, 433)	
LS Mean (SE)	241.2 (6.19)	222.9 (6.30)	0.0341 <sup>a</sup>
95% CI of LS Mean	229.1, 253.4	210.5, 235.3	
			0.7276 <sup>b</sup>
<b>Treatment Week 1 (n)</b>	150	143	
Mean (SE)	270.8 (6.03)	238.5 (7.42)	
Median (min, max)	265.4 (91, 480)	240.0 (42, 420)	
LS Mean (SE)	264.2 (4.74)	247.2 (4.85)	0.0107 <sup>c</sup>
95% CI of LS Mean	254.9, 273.6	237.6, 256.7	
			0.1592 <sup>d</sup>
<b>Treatment Week 2 (n)</b>	140	139	
Mean (SE)	270.1 (5.98)	243.3 (7.38)	
Median (min, max)	274.6 (57, 540)	240.0 (0, 450)	
LS Mean (SE)	263.9 (4.88)	250.4 (4.87)	0.0469 <sup>c</sup>
95% CI of LS Mean	254.3, 273.5	240.8, 260.0	
			0.9384 <sup>d</sup>

**Table 10** Study ZI-12: Sponsor's Analysis of Subjective Total Sleep Time (continued)

	<b>Zolpidem 3.5 mg</b>	<b>Placebo</b>	<b>P-value</b>
<b>Treatment Week 3 (n)</b>	137	134	
Mean (SE)	270.7 (6.66)	245.0 (7.97)	
Median (min, max)	270.0 (44, 540)	240.0 (30, 480)	
LS Mean (SE)	263.8 (5.64)	254.6 (5.69)	0.2452 <sup>c</sup>
95% CI of LS Mean	252.6, 274.9	243.4, 265.8	
			0.7685 <sup>d</sup>
<b>Treatment Week 4 (n)</b>	129	130	
Mean (SE)	272.8 (7.26)	250.0 (7.82)	
Median (min, max)	270.0 (68, 540)	254.2 (58, 437)	
LS Mean (SE)	266.1 (5.46)	260.0 (5.43)	0.4233 <sup>c</sup>
95% CI of LS Mean	255.3, 276.8	249.3, 270.7	
			0.9235 <sup>d</sup>
<b>DB Treatment (n)</b>	150	144	
Mean (SE)	270.7 (5.63)	246.9 (7.18)	
Median (min, max)	270.0 (70, 450)	247.8 (55, 420)	
LS Mean (SE)	264.1 (4.25)	255.0 (4.33)	0.1282 <sup>c</sup>
95% CI of LS Mean	255.7, 272.4	246.5, 263.5	
			0.4154 <sup>d</sup>

\*Baseline sTSTMOTN = average of sTSTMOTN values collected during the 2-week placebo single blind screening period for those nights study medication was taken.

a. The 2-way ANOVA with mean sTSTMOTN as the response; treatment and pooled site as fixed effects. P-value shown is for treatment.

bP-value for pooled site-by-treatment interaction when pooled site-by-treatment interaction was added to the model given in a.

cANCOVA model with mean sTSTMOTN as the response; treatment and pooled site as fixed effects and average sTSTMOTN at baseline as a covariate. P-value shown is for treatment.

dP-value for pooled site-by-treatment interaction when pooled site-by-treatment interaction was added to the model given in c.

### **sNAWMOTN Post-middle-of-the-night Awakening (Sponsor's Analysis)**

Subjective number of awakenings after falling back to sleep were based on the recorded response to the following IVRS question, "After you fell back to sleep, how many times did you wake up again before waking up in the morning?" Results of the secondary end point, sNAWMOTN, are presented in Table 11.

Due to the non-normal distribution of the observations for this parameter, sNAWMOTN was subjected to a post-hoc categorical analysis. Actual subject values of sNAW for a single night

would be integer values (no awakenings, 1 awakening, etc), so categories were divided at the integer values of 0, 1, and 2. Table 11 presents the categorical analysis results for sNAWMOTN. There was no significant difference between groups at baseline. Statistically significant differences in favor of zolpidem tartrate sublingual lozenge versus placebo for sNAWMOTN were seen across the double-blind treatment period ( $p < 0.001$ ) as well as at weeks 1 through 3 of the study. Treatment effect was evident as higher percentages of subjects with 0 awakenings were in the zolpidem tartrate sublingual lozenge group versus the placebo group during the entire 4 weeks of double-blind treatment.

Table 11 Study ZI-12: Sponsor's Analysis of Number of Awakenings after Treatment

	3.5 mg Zolpidem	Placebo	P-value <sup>a</sup>
Average Number of Awakenings Reported	% patients reporting		
<b>Baseline Average*</b>			0.160
0 awakenings	16.7%	11.8%	
>0 and ≤1 awakening	46.0%	45.1%	
>1 and ≤2 awakenings	26.0%	27.1%	
>2 awakenings	11.3%	16.0%	
<b>Week 1</b>			0.027
0 awakenings	40.7%	30.6%	
>0 and ≤1 awakening	32.7%	34.7%	
>1 and ≤2 awakenings	18.7%	16.7%	
>2 awakenings	8.0%	18.1%	
<b>Week 2</b>			0.011
0 Awakenings	43.3%	30.9%	
>0 and ≤1 awakening	33.3%	35.3%	
>1 and ≤2 awakenings	17.7%	20.1%	
> 2 Awakenings	5.7%	13.7%	
<b>Week 3</b>			0.007
0 Awakenings	48.9%	34.3%	
>0 and ≤1 awakening	27.0%	31.3%	
>1 and ≤2 awakenings	16.8%	17.2%	
> 2 Awakenings	7.3%	17.2%	
<b>Week 4</b>			0.064
0 Awakenings	45.0%	35.4%	
>0 and ≤1 awakening	29.5%	30.8%	
>1 and ≤2 awakenings	20.2%	23.1%	
> 2 Awakenings	5.4%	10.8%	

**Table 11 Study ZI-12: Sponsor's Analysis of Number of Awakenings after Treatment (continued)**

	<b>3.5 mg Zolpidem</b>	<b>Placebo</b>	<b>P-value<sup>a</sup></b>
<b>Average Number of Awakenings Reported</b>	<b>% patients reporting</b>		
<b>DB Treatment</b>			<0.001
No Awakenings	30.0%	16.7%	
>0 and ≤1 awakening	46.0%	43.1%	
>1 and ≤2 awakenings	17.3%	25.0%	
> 2 Awakenings	6.7%	15.3%	

\*Baseline sNAWMOTN=average of sNAWMOTN values collected during the 2-week placebo single blind screening period for those nights study medication was taken.

Note: Continuous values obtained by averaging were then categorized as shown.

a The row mean score statistic for categorized mean sNAWMOTN; p-value shown is for treatment.

### **sWASOMOTN Post-middle-of-the-night Awakening**

sWASOMOTN was defined as the number of minutes recorded in response to the following IVRS question: "Considering all of these awakenings (after taking study medication and returning to sleep), how long were you awake from the time you went back to sleep after dosing until you got out of bed this morning?" Subjective wake time is presented in Table 12. Due to the non-normal distribution of the observations for this parameter, sWASOMOTN was subjected to a post-hoc categorical analysis. There was no significant difference between groups at baseline. Nominally significant differences favored zolpidem tartrate sublingual lozenge versus placebo across the 4-week double-blind treatment ( $p=0.006$ ) as well as at weeks 1 through 3 of the study. The treatment effect was observed on Week 1 with a steady increase in the number of subjects with no sWASOMOTN corresponding to the length of treatment.

Table 12 Study ZI-12: Sponsor's Post-Hoc Analysis of Subjective WASO

Average Wake Time After Sleep Onset	3.5 mg Zolpidem	Placebo	Overall P-value <sup>a</sup>
	% Patients Reporting		
<b>Baseline*</b>			0.512
No Wake Time	16.7%	11.8%	
>0-20 Minutes	18.7%	14.6%	
21-60 Minutes	25.3%	32.6%	
> 60 Minutes	39.3%	41.0%	
<b>Week 1</b>			0.021
No Wake Time	40.7%	30.6%	
>0-20 Minutes	16.7%	14.6%	
21-60 Minutes	22.0%	24.3%	0.011
> 60 Minutes	20.7%	30.6%	
<b>Week 2</b>			0.005
No Wake Time	43.3%	30.9%	
>0-20 Minutes	24.8%	17.3%	
21-60 Minutes	14.2%	26.6%	
> 60 Minutes	17.7%	25.2%	
<b>Week 3</b>			0.042
No Wake Time	48.9%	34.3%	
>0-20 Minutes	13.1%	17.9%	
21-60 Minutes	16.8%	23.1%	
> 60 Minutes	21.2%	24.6%	
<b>Week 4</b>			0.063
No Wake Time	45.0%	35.4%	
>0-20 Minutes	19.4%	18.5%	
21-60 Minutes	17.1%	23.1%	
> 60 Minutes	18.6%	23.1%	
<b>DB Treatment</b>			0.006
No Wake Time	30.0%	16.7%	
>0-20 Minutes	27.3%	26.4%	
21-60 Minutes	26.0%	31.9%	
> 60 Minutes	16.7%	25.0%	

\*Baseline sWASOMOTN = average of sWASOMOTN values collected during the 2-week placebo single-blind run-in period for those nights study medication was taken. Note: Continuous values obtained by averaging were then categorized as shown.

<sup>a</sup>The row mean scores statistic for categorized mean sWASOMOTN; p-value shown is for treatment.



### 3.1.2.5 Reviewer's Results

#### 3.1.2.5.1 Primary Analysis

The primary analysis averaged the Latency to Sleep Onset over the nights on which the subject had an awakening, took the study drug, and provided efficacy data for that night. During the 14 day baseline phase patients in both groups had on average 9.7(ranging from 2 to 14) nights of dosing. During the 28 day double blind treatment phase patients assigned to Zolpidem had on average 17.3(ranging from 1 to 28) nights of dosing as compared to 17.8 (ranging from 1 to 28) for placebo. The average time of first dose was day 1.6 (range: 1 to 7) for the Zolpidem group and day 1.4 for placebo (range: 1 to 7). The average time of last dose was day 25 for both groups (ranging from 1 to 28). The average LSO prior to the MOTN awakening was comparable between the groups (51.5 vs. 49.1,  $p=0.552$ ). This reviewer verified the sponsor's primary analysis result for the latency to sleep onset post middle of the night awakening which found that Zolpidem was statistically significantly lower than placebo (38.2 vs. 56.4,  $p<0.0001$ ).

#### 3.1.2.5.2 Assessment of the Impact of Missing Data on the Primary Analysis Result

Fourteen Zolpidem group patients (9.2%) and nine (6.2%) Placebo group patients did not complete the study. The means presented below are LS means based on the prespecified analysis model. They have been transformed from the log scale back to the original scale by means of the exponential function,  $e^x$ . The primary analysis seemed to be relatively robust to different ways of handling missing data. In particular, if we analyze the LSO post MOTN awakening on the first night the patient dosed (LS Means 42.4 for Z vs. 54.4 for Pl,  $p=0.0009$ ), the last night the patient dosed (34.1 vs. 52.2,  $p<0.0001$ ) or any particular one of the four weeks then the result remains significant at the nominal level. In the subgroup of completers the treatment group difference was also significant (37.6 vs. 56.8,  $p<0.0001$ ). In the subgroup of those who dosed in the last scheduled week ( $N=129$  for Z, 130 for Pl) the result was also nominally significant, (32.6 vs. 50.9,  $p<0.0001$ ). Finally, if for those who dropped out before the last week we average the available data with an imputation for the last week based on a LSO post MOTN value of 130 (beyond the 95% of the observed distribution) for the Zolpidem group and 0 for the placebo group the result is still nominally significant (41.6 vs. 55.1,  $p<0.0001$ ). Considering these analyses just described, the primary analysis result seems robust to varying assumptions regarding the missing data.

#### 3.1.2.5.3 Analyses of Key Secondary Endpoints

There was a bit of an imbalance at baseline in TST: 241.6 for Zolpidem and 223.3 for placebo,  $p=0.035$ . However, if we consider that many variables were compared at baseline then after adjusting for multiplicity this would not be statistically significant and the analysis of TST was adjusted for baseline TST (baseline TST was included as a covariate). The statistical test should still be valid despite the possible baseline imbalance and it did not find a statistically significant difference between the treatment groups in terms of TST after MOTN awakening,  $p=0.128$ .

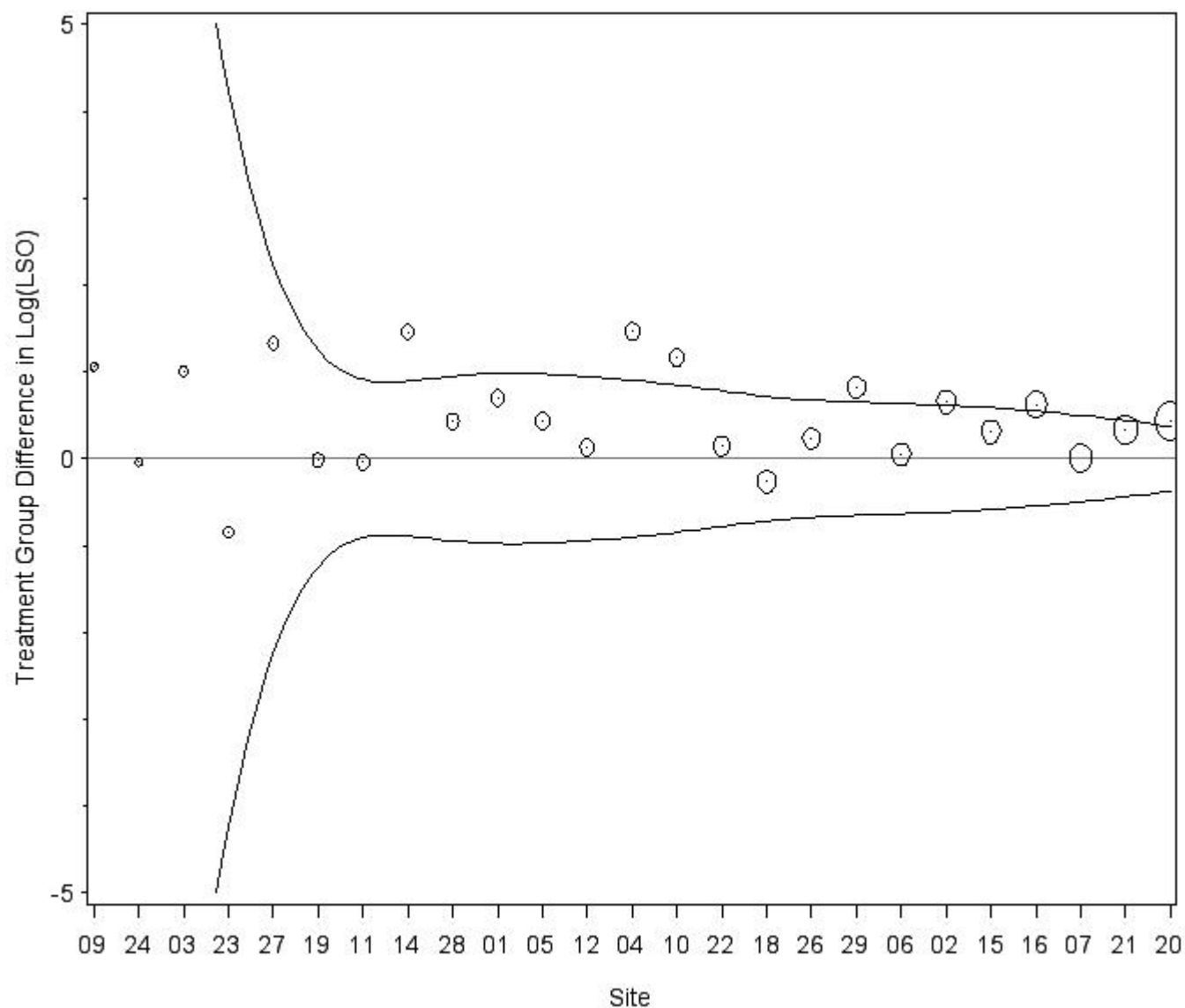
The second endpoint specified in the hierarchy of key secondaries was the number of awakenings post MOTN. Count regression methods adjusted for the average number of post MOTN awakenings over the baseline period do not support the sponsor's post-hoc categorical analysis of the number of awakenings post MOTN awakening. The count regression model used by this reviewer incorporated a random effect for patient to account for the fact that a patient's number of awakenings on a given night is likely positively correlated with the number from the previous night. Count regression methods are often used or proposed for this type of data because the underlying Poisson distribution (or, alternatively, the Negative Binomial distribution) for the number of awakenings post MOTN on a given night only has positive probability at zero and positive integer values (0, 1, 2, 3, ..., etc. ) which fits this data since fractional awakenings are not possible. This is in contrast to the normal distribution which models the data as if fractional awakenings are possible. These count regression analyses cast some doubt on the sponsor's post hoc analyses of the number of awakenings post MOTN. Also, perhaps more importantly, TST was higher in the hierarchical order of secondary endpoints and the treatment effect on it was not statistically significant.

The last endpoint in the prespecified hierarchy of key secondaries was the subjective WASO post MOTN endpoint. Estimated mean subjective WASO post MOTN was 38.8 for placebo and 32.5 for 3.5 mg,  $p=0.0857$ . The sponsor argues against this analysis claiming that WASO was not normally distributed which has some merit but they did not prespecify any alternative analysis. Therefore, it is difficult to interpret a p-value other than the one corresponding to the pre-planned analysis for the WASO post MOTN. In addition, as mentioned above the result on the TST, the first endpoint in the hierarchy, was not significant.

#### ***3.1.2.5.4 Observed Treatment Effects by Site***

There were 25 sites which ranged in size from 3 to 64 randomized patients. The treatment group difference favored Zolpidem numerically in 20 of the 25 sites. Excluding any one particular site did not alter the significance of the result for LSO post MOTN awakening. The observed treatment differences are shown by site in Figure 3. The size of the circle is proportional to the number of patients in the site and the curves on the plot correspond to the critical t-statistic value (based on a 0.05 significance level) for the corresponding sample size.

**Figure 3 Study ZI-12 Treatment Effect on Log(Latency to Sleep Onset) by Site**



### 3.2 Evaluation of Safety

Safety was not reviewed in this document. Please see the medical officer's review for the review of safety.

## 4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

### 4.1 Gender, Race and Age

#### 4.1.1 Gender

In study ZI-06-010 there was very little basis to compare genders with regards to efficacy because only 24 (29%) of 82 were males. The complex crossover design further complicates this assessment. Based on all available data the interaction between treatment group and gender was not significant,  $p=0.33$ . Based on the first period data only, the interaction between treatment group and gender was also not significant,  $p=0.77$ . Therefore, based on what limited data there is for comparing genders in study ZI-06-010 there is no evidence of a different treatment effect depending on the gender. The LS Mean after transforming it back from the log scale to the original scale was 29.6 for females on placebo, 11.2 for females on 3.5 mg, 32.8 for males on placebo, and 13.6 for males on 3.5 mg based on period 1 data only.

In study ZI-12 the majority of patients were female: 203 (68%) of 297 the randomized patients with post-baseline efficacy data were female. There was no compelling evidence that the treatment effect on LSOMOTN depended on gender,  $p=0.7682$ . This is relatively clear from the model incorporating a treatment by gender interaction from which the LS Means were estimated as 37.5 for Zolpidem 3.5 mg Females, 55.9 for Placebo Females, 40.0 for Zolpidem 3.5 mg Males, and 57.2 for Placebo Males.

#### 4.1.2 Race

In Study ZI-06-010 the distribution of race was: 42 (51%) of 82 were classified as Caucasian, 36(44%) of 82 were classified as Black, and only 5% were classified as other (Asian or Hispanic). The complex crossover design further complicates the comparison of the races with respect to efficacy. Based on all available data the interaction between treatment group and race on LPS post MOTN was not significant,  $p=0.41$ . Based on the first period only the interaction between treatment group and race was also not significant,  $p=0.69$ . Therefore, based on what limited data there is for comparing races in study ZI-06-010 there is no evidence of a different treatment effect depending on the race. The LS Mean after transforming it back from the log scale to the original scale was 29.4 for Caucasians on placebo, 12.7 for Caucasians on 3.5 mg, 32.7 for Blacks on placebo, and 12.2 for Blacks on 3.5 mg in period 1, 29.6 for Others on placebo, and 5.7 for Others on 3.5 mg.

In study ZI-12 the distribution of race was: 65% were classified as White, 31% were classified as Black or African American and 4% were Other (American Indian, Asian, Hawaiian, or unspecified Other). In study ZI-12 there was not compelling evidence that the treatment group difference on Latency (LSO) post MOTN depended on Race,  $p=0.110$ . Based on the model with interaction the estimated LS Means transformed back to the original scale from the log scale were:

35.2 for Zolpidem vs. 45.9 for placebo in Whites  
 37.8 for Zolpidem vs. 55.4 for placebo in African Americans  
 56.9 for Zolpidem vs. 76.3 for placebo in Others.

Differences between the race groups with the most data were slight and evidence for an interaction between race and treatment group is relatively weak.

### 4.1.3 Age

In study ZI-10 the mean age was 45.9 (range: 19 to 64). In study ZI-06-010 there was a significant interaction between age and treatment group on LPS after MOTN awakening,  $p=0.0498$ , based on data from all three periods. This suggests that the treatment effect increased slightly with age. In particular the difference on the log scale is estimated to be 0.947 (17.9 minutes) at age 45, 1.136 (21.3 minutes) at age 55, and 1.325 (24.7 minutes) at age 65, and 1.513 (28.1 minutes) at age 75. However, this effect is debatable because there is no evidence of an interaction between age and treatment group if we only look at the first period data,  $p=0.76$ .

In study ZI-12 the mean (as well as median) age was 43. Ages ranged from 18 to 64. There was again the suggestion that the treatment effect may increase with age,  $p=0.0136$ . This model suggests that the difference increases linearly on the log scale following the equation  $.191 - .0136 \times \text{Age}$ . This equation evaluates to a negative value for all ages observed in the trial with negative differences meaning the average LSO post MOTN is smaller for Zolpidem. Estimated LS Means transformed back to the original scale from the log scale were for example: 43.2 vs. 53.7 at age 30, 39.1 vs. 55.8 at age 40, 35.5 vs. 58.0 at age 50, 32.1 vs. 60.2 at age 60 for 3.5 mg and placebo, respectively. The sponsor investigated the age groups  $<55$ , age  $\geq 55$ . It was less clear that the treatment effect depended on age when ages were grouped this way,  $p=0.135$ , but numerically at least, the estimates based on the analysis of these age groups suggested the same trend for increasing effect with increasing age group.

This possible age by treatment group interaction may not be a big issue as it was quantitative, i.e., the sign of the estimated treatment group difference favored Zolpidem for all ages studied.

## 4.2 Other Special/Subgroup Populations

No other Special/Subgroups Populations were examined.

## 5 SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

In study ZI-06-010 the Latency to Persistent Sleep before scheduled MOTN awakening was measured each study night. There was some suggestion in the data that the LPS pre MOTN awakening varied significantly with the crossover period as well as the treatment group, despite it corresponding to the period before drug administration. However, it was prespecified that the primary analysis of the primary endpoint, LPS post MOTN awakening, would use the Total Sleep Time before MOTN awakening as the covariate instead of using the LPS pre MOTN awakening as the covariate. Furthermore, there was no period or treatment effect apparent on the TST before MOTN awakening. The primary analysis found that Zolpidem was superior to placebo in reducing Latency to Persistent Sleep as measured by PSG. In spite of this possible group difference in LPS prior to scheduled awakening the pre-specified primary analysis as well as most other analyses (e.g., restricted to the first of two days in each crossover period or restricted to a particular crossover period) seem to support a significant treatment effect on LPS after scheduled awakening. Therefore, in light of the balance of the data suggesting a robustness of the statistical significance of the treatment effect we may be inclined to disregard the observed group differences in pre-MOTN LPS. Note also that the primary analysis of LPS post MOTN prespecified adjusting for total sleep time before scheduled awakening rather than LPS before scheduled awakening and there were no significant group imbalances in total sleep time before scheduled awakening.

The analysis plan stated that following secondary endpoints based on the time after the scheduled middle of the night awakening would be tested in this order: TST based on PSG, Sleep efficiency based on PSG, Subjective Sleep quality rating, Subjective Sleep Onset Latency, Subjective TST, and Latency to Persistent Sleep based on PSG. The last one in the list is for the low dose vs. placebo comparison and the rest are for the high dose vs. placebo comparison. This reviewer verified that the prespecified analysis for each of these secondary endpoints was significant at the 0.05 level. Note that Sleep Efficiency endpoint is proportional to  $TST / (Time\ in\ Bed)$ . In this trial the Pearson correlation between TST and sleep efficiency is estimated by this reviewer to be 0.999. Therefore, in this reviewer's opinion the Sleep Efficiency endpoint does not seem to provide any distinct information from the TST endpoint.

In study ZI-12 there was a bit of an imbalance at baseline in TST post MOTN awakening, which was the highest secondary endpoint in the hierarchy: 241.6 for Zolpidem and 223.3 for placebo,  $p=0.035$ . However, if we consider that many variables were compared at baseline then after adjusting for multiplicity this would not be statistically significant and the analysis of TST was adjusted for baseline TST (baseline TST was included as a covariate). The statistical test should still be valid despite the possible baseline imbalance and it did not find a statistically significant difference between the treatment groups in terms of TST after MOTN awakening,  $p=0.128$ . Since the TST was the first secondary endpoint in the prespecified testing hierarchy and the test was not statistically significant the results for the subjective WASO and subjective number of

awakenings should be considered exploratory in accordance with the prespecified multiplicity adjustment plan.

There was a slight suggestion in both studies that the treatment effect on Latency increased somewhat with age. However, if we focused on the first period of the crossover study then no such trend was apparent. Similarly, in study ZI-12, if we compared the treatment group difference in those Age < 55 vs. those Age ≥ 55, age groups defined by the sponsor, then the difference was not significantly different between these age groups. Therefore, the interaction effect is debatable and in any case the treatment group difference favored Zolpidem on average numerically for all ages studied.

## **5.2 Conclusions and Recommendations**

The data from the key efficacy studies, ZI-06-010 and ZI-12, seem to support the superiority of the Zolpidem tartrate sublingual lozenge over placebo for reducing latency to persistent sleep after a middle of the night awakening.

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

/s/

-----  
Tristan Massie  
6/19/2009 10:19:01 AM  
BIOMETRICS

Kun Jin  
6/19/2009 10:40:20 AM  
BIOMETRICS

Kooros Mahjoob  
6/19/2009 02:04:24 PM  
BIOMETRICS