

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
022036Orig1s000

STATISTICAL REVIEW(S)

NDA: 22036 (S0019)

Drug: Doxepin

Indication: Insomnia

Type of Submission: Complete Response Submission

Date of Submission: June 4, 2009

Medical Officer: June Cai, M.D.

Statistical Reviewer: Tristan Massie, Ph.D.

1. Executive Summary

1.1. Conclusions and Recommendations

The sponsor has done post-hoc analyses to dispute the apparent lack of consistency over time of the treatment differences in terms of subjective WASO and subjective TST. These analyses make an assumption of constancy over time that does not seem to be justified and without this assumption they essentially confirm the original protocol specified analyses. Therefore, for subjective WASO and TST we are left with the inconsistency of treatment effects over time as determined by the original protocol specified analyses. In particular, there is not solid evidence for a treatment difference in subjective WASO (or TST) between 6mg and placebo at the end of the non-elderly adult study.

1.2. Statistical Issues for Doxepin Complete Response Submission

The sponsor has submitted post-hoc analyses to dispute the inconsistency of the treatment effects over time on subjective maintenance. Please refer to my statistical review of the original NDA submission for further details of the studies and the prespecified analyses. They argue in the new submission that the treatment by time interaction is not statistically significant based on a MMRM (mixed model for repeated measures) analysis so that the average treatment difference over the entire double blind period can be used to represent the difference at the end of the double blind period. However, the non elderly adult study (501) was not powered for a test of the treatment by time interaction (estimated power based on simulation was 43%) and failing to reject the hypothesis of no interaction only means there is not enough evidence to reject the hypothesis of no interaction it does not imply that this null hypothesis is true. In order to have 80% power at the 0.05 significance level for the treatment by time interaction test we would need about twice as many patients (170) per group to detect an interaction of a size corresponding to the interaction parameter estimates based on this data. In other terms, with the given sample size, to have 80% power for the interaction test the size of the difference between the treatment effect at day 16 and day 30 would have to be about 26 minutes. To better answer the question of whether any interaction exists one needs to take an equivalence testing approach, i.e., let the null hypothesis be that two or more of the time specific treatment group differences between Doxepin 6 mg and placebo differ by, for example, 10 or more minutes and the alternative hypothesis be that all their time specific treatment differences are consistent to within 10 minutes. This reviewer found that an equivalence testing approach to the treatment by time interaction suggests that we can not conclude that all of the time specific treatment differences in terms of sWASO (subjective Wake

Time after Sleep Onset) are consistent to within 10 minutes. This seems to be an important difference considering that it is about 50% of the estimated treatment difference on Night 1. Also, the assessment schedule probably didn't have enough coverage of the 30 day range to detect a non-constant relationship, i.e., we can't necessarily assume that the treatment difference is constant at all times in between scheduled assessment times. If it is not then it wouldn't make much sense to average over the three assessments to estimate the effect at the end, whether or not the effect is constant at these three times. For these reasons, to correctly assess the treatment difference at the last visit we need to focus on the data from that visit alone rather than averaging over the entire double blind period. This means that we are left with the inconsistency of treatment effects over time as determined by the original protocol specified analyses.

In study 501, the non-elderly adult study, the sponsor prespecified a preference for the subjective TST over the subjective WASO in the original analysis plan: "while the accepted measure of objective sleep maintenance is WASO, subjective total sleep time (sTST) is the preferred measure of subjective sleep maintenance". The subjective TST results for the treatment difference at the final visit are not nominally significant regardless of whether we look at the first night sTST or the sTST averaged over the two nights of the visit ($p=0.4706$ and 0.1464 , respectively). On the other hand, the subjective WASO results for the last visit are nominally significant if we look at the mean of the 2 nights ($p=0.0199$) instead of the first night only ($p=0.6282$, prespecified timepoint for analysis).

Study 503 results for the 3 mg vs. placebo difference in subjective maintenance were also not consistent over time. As for study 501 while the treatment by time interaction did not reach statistical significance at 0.05 or 0.10 this reviewer found that using an equivalence testing approach we can not conclude that all of the time-specific treatment differences are within 10 minutes of each other. Therefore, to correctly assess the treatment difference at the last visit it seems that we need to focus on the data from that visit alone rather than averaging over the entire double blind period. This means that we are left with the inconsistency of treatment effects over time as determined by the original protocol specified analyses.

The IVRS measured subjective TST was an exploratory endpoint. The usefulness of this data is limited by its exploratory designation as well as the facts that a substantial proportion of randomized patients (34%) did not have a baseline assessment and there is a statistically significant group difference between 6 mg and placebo at baseline among those that had baseline assessments ($p=0.03$). This latter fact means that any post-baseline group difference might be due to the baseline difference rather than the treatment. Also, because this was a study in the elderly it does not directly address subjective maintenance for non-elderly adults where the evidence is most lacking.

2. Introduction

The Complete Response Letter outlined the Division's concerns with respect to the safety and efficacy of Silenor after review of the original NDA. To obtain clarity regarding requirements for resubmission, Somaxon met with the Division at an End-of-Review Meeting on April 6, 2009 (see minutes for the End-of-Review Meeting, 04 May 2009).

The Division's concerns relating to efficacy, as stated in the Complete Response Letter and clarified at the End-of-Review Meeting, were summarized by the sponsor in the complete response submission as follows.

- The Division noted that results for subjective WASO (sWASO) in adults were not significant at Night 15 (Day 16) and Night 29 (Day 30) for Study SP-0501. However, they noted that nominally significant effects (unadjusted for multiplicity) for doxepin 6 mg were seen at related times not specified for analysis in the analysis plan: Night 16 (Day 17) and Night 30 (Day 31), as well as for the average of Nights 15 and 16 (Days 16 and 17) and Nights 29 and 30 (Days 30 and 31).
- The Division noted that results for sWASO in elderly were not significant at Night 1 (Day 2) for Study SP-0503. However, they noted that doxepin 3 mg showed nominal significance compared to placebo for Night 29 (Day 30) and Night 85 (Day 86).

3. Details of the Complete Response Submission

The sponsor's summary of clinical efficacy addendum includes supplemental post-hoc analyses for subjective sleep maintenance data from SP-0501 and SP-0503 using a mixed-effect model repeated measures (MMRM) approach (see [Hedeker and Gibbons, 2006](#), for a detailed overview). For these studies, both sTST and sWASO (obtained from the morning questionnaire) were analyzed using the MMRM method (separate models for each study and outcome variable) that included fixed effects for treatment group, time (as a discrete factor), the treatment-by-time interaction, and the baseline value of the endpoint. In order to avoid the potential sensitivity of tests and estimators to the choice of an arbitrary covariance structure, each analysis used an unstructured covariance matrix for the repeated observations within each subject (Davis, 2002, chapter 6).

Due to the schedule of timepoints at which repeated measurements were obtained in Study SP-0501, the time periods for the repeated measurements were specified using two methods:

- Using the study visit as the time variable. In this case, the results from each pair of adjacent days were averaged (Visit 4 [Days 2 and 3], Visit 5 [Days 16 and 17], Visit 6 [Days 30 and 31]) prior to analysis. This method leads to a maximum of three repeated measurements per subject.
- Using each study day as the time period variable. Here, the result from each day was treated as a unique repeated measurement. This method leads to a maximum of six repeated measurements per subject.

The degree of sleep disturbance in those with insomnia can be quite variable from night to night, and may include nights without any sleep disturbance (Edinger et al, 2004). In the

sleep laboratory setting it is common to use the average of a pair of nights to assess efficacy (Roth et al, 2006a; Roth et al, 2006b; Scharf et al, 1994). The sponsor states that because the best estimate for sleep is obtained by averaging over paired nights, the primary MMRM analysis for study SP-0501 will use study visit as the time variable, such that the data from each pair of adjacent days will be averaged prior to analysis. The MMRM analysis using each study day as the time variable is considered to be secondary for SP-0501. Note that the prespecified analysis did not average over paired nights but rather only used the first night of each pair.

Although these analyses were conducted after the submission of the original NDA, a pre-specified analysis plan was followed. According to that plan, the first step was to test the significance of the treatment-by-time interaction. A p-value of greater than 0.1 ($p > 0.1$) was prespecified as the criterion for concluding that the effects of treatment did not differ significantly over time. In addition, the two components of the treatment-by-time interaction (for each dose versus placebo separately) were also tested.

If the p-value from the test of treatment-by-time interaction was not significant ($p > 0.1$), then the overall treatment differences (each dose versus placebo separately) were estimated by the contrast averaging the corresponding treatment differences at each time point. The null hypothesis that the treatment difference is zero was also tested. In the absence of a significant treatment-by-time interaction, the most relevant statistical test and associated estimate of treatment effect corresponds to that of the overall treatment difference. However, for completeness of presentation, the estimates of the overall estimate of treatment effect are displayed along with estimates of treatment effect at each time point, regardless of the conclusion concerning the treatment-by-time interaction.

Reviewer's Comment:

The sponsor's assertion that the new analyses were pre-specified has little relevance as the plans were developed well after the unblinding of the data, the completion of the original study report, and the FDA decision on the original application. In addition, this reviewer does not believe that the new analysis plan was submitted to FDA for review.

In the Complete Response Letter, the Division stated:

“In order for a hypnotic drug product to be approved, its effect on both objective and subjective measures of particular sleep difficulties must be established. Further, it is expected that any treatment for patients with chronic insomnia will be shown to be effective not only at the beginning of treatment, but also that its effects will persist out in time (at least for one month).”

In the complete response submission the sponsor argues that a mixed model analysis approach (MMRM) is more appropriate than the pre-specified ANCOVA analysis. This would be arguable before the study was unblinded but not after. Here, in the absence of any major shortcomings of the prespecified ANCOVA that are addressed by the MMRM the ANCOVA should still be considered the primary analysis.

4. Study 501

4.1. Sponsor's Post-hoc analyses

The fact that separate assessments were made on 2 days during each visit creates a multiplicity type I error control issue if we consider straying from the protocol specified approach of only using the first day of each visit for the analysis.

The LS mean differences from placebo for sWASO as determined by the sponsor for the first and last visit are presented in Table 1.

Table 1 Sponsor's LS Mean Difference from Placebo (Minutes) for sWASO: ITT Analysis Set

Study	Time point	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
0401	Mean Day 2 & Day 3	2.5	-2.7	-8.3
0402	Mean Day 2 & Day 3	-14.5*	-18.5**	-19.1**
0501	Day 2	--	-20.6***	-20.3***
	Mean Day 2 & Day 3	--	-13.7**	-18.0***
	Day 30	--	3.2	-3.4
	Mean Day 30 & Day 31	--	-5.8	-14.2*
	Overall (MMRM) ¹	--	-10.2*	-14.2**
	Overall (ANCOVA)	--	-10.0*	-14.2***
0503	Day 2	-1.9	-19.2	--
	Day 86	-29.7**	-24.5*	--
	Overall (MMRM) ²	-10.1	-18.3**	--
	Overall (ANCOVA)	-9.2	-16.9**	--
0509	Week 1	--	--	-22.4***
	Week 4	--	--	-19.5***
0502	Day 2	--	--	-10.6**

*p≤0.05; **p≤0.01; ***p≤0.001; all such noted results favor doxepin over placebo.

1. For supplemental MMRM analysis in **SP-0501**, p-values for treatment-by-time interactions were 0.7411 (overall), 0.5943 (doxepin 3 mg), and 0.5612 (doxepin 6 mg). According to the prespecified analysis plan for supplemental analyses, overall effects of treatment are therefore presented.
2. For supplemental MMRM analysis in **SP-0503**, p-values for treatment-by-time interactions were 0.2816 (overall), 0.0891 (doxepin 1 mg), and 0.6254 (doxepin 3 mg). According to the prespecified analysis plan for supplemental analyses, overall effects of treatment are therefore presented. As Somaxon is no longer seeking approval for doxepin 1 mg, overall results are presented for completeness despite the lower p-value associated with the interaction term for 1 mg.

The LS mean differences from placebo for sTST as determined by the sponsor for the first and last visit are presented in Table 2.

Table 2 Sponsor’s LS Mean Difference from Placebo (Minutes) for sTST: ITT Analysis Set

Study	Time point	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
0401	Mean Day 2 & Day 3	-0.7	15.5	16.2*
0402	Mean Day 2 & Day 3	16.3*	23.3***	31.1***
0501	Day 2	--	23.3**	22.0*
	Mean Day 2 & Day 3	--	16.1*	20.5**
	Day 30	--	0.2	9.7
	Mean Day 30 & Day 31	--	7.1	12.3
	Overall (MMRM) ¹	--	11.9*	17.3**
	Overall (ANCOVA)	--	10.8	16.6**
0503	Day 2	-4.0	19.4	--
	Day 86	27.3**	22.5*	--
	Overall (MMRM) ²	9.5	18.9*	--
	Overall (ANCOVA)	10.5	18.5*	--
0509	Week 1	--	--	28.8***
	Week 4	--	--	24.9***
0502	Day 2	--	--	32.9***

*p≤0.05; **p≤0.01; ***p≤0.001; all such noted results favor doxepin over placebo.

1. For supplemental MMRM analysis in [SP-0501](#), p-values for treatment-by-time interactions were 0.9609 (overall), 0.8151 (doxepin 3 mg), and 0.8050 (doxepin 6 mg). According to the prespecified analysis plan for supplemental analyses, overall effects of treatment are therefore presented.
2. For supplemental MMRM analysis in [SP-0503](#), p-values for treatment-by-time interactions were 0.2046 (overall), 0.1175 (doxepin 1 mg), and 0.5720 (doxepin 3 mg). According to the prespecified analysis plan for supplemental analyses, overall effects of treatment are therefore presented.

As Somaxon and the Division have noted, some individual timepoints do not show a statistically significant effect of Silenor relative to placebo; however, Somaxon contends that using the MMRM method to analyze and interpret data in a longitudinal study is more efficient than point-wise analysis ([Mallinckrodt et al, 2008](#)). In their opinion in the absence of any indication of a treatment-by-time interaction, the most reliable estimates of the treatment effect come from the longitudinal models, not from point-wise estimates. They believe the results from the longitudinal analyses clearly show that doxepin 3 mg and 6 mg are effective on subjective measures of sleep maintenance when compared to placebo. However, these MMRM analyses are entirely post-hoc. They don’t seem to add much other than the average effect over the double blind period but the relevance of that hinges on the assumption that there is no treatment by time interaction. Furthermore, if we are primarily interested in whether there is an effect at the end of the treatment period then the first visit should not be included in the average as it is in the sponsor’s analysis.

Table 3 presents the sponsor’s results for sWASO when the average over the 2 nights of each visit is the dependent variable. Note that the prespecified analysis plan called for using only the first night of each visit in the analysis, so these analyses are post-hoc. The “Overall” timepoint represents the result for the treatment difference averaged over the three visits.

Table 3 Study 501: sWASO averaged over 2 nights of each Visit

Timepoint	Doxepin 3 mg		Doxepin 6 mg	
	LS Mean Difference from Placebo (Std. Err)	p-value	LS Mean Difference from Placebo (Std. Err)	p-value
Overall	-10.2 (4.41)	0.0213	-14.2 (4.41)	0.0014
By Visit				
Visit 4 (Days 2 & 3)	-13.3 (4.84)	0.0065	-17.8 (4.87)	0.0003
Visit 5 (Days 16 & 17)	-10.4 (5.92)	0.0804	-11.4 (5.90)	0.0539
Visit 6 (Days 30 & 31)	-7.0 (6.23)	0.2656	-13.5 (6.23)	0.0316

Notes: Subjects provided sWASO values on both days of the study visit; data from paired days were averaged prior to analysis. See SCE [Addendum Attachment 1 Section 6.1](#) for an overview of the extent of missing data by study visit. Values of sWASO >8 hours are excluded from the analysis prior to averaging across paired days (one value met this criterion; see SCE [Addendum Attachment 1 Table A3](#) for details). P-values comparing each active treatment versus placebo are obtained from a MMRM analysis that includes the main effects for treatment, visit, and the treatment-by-visit interaction with the baseline value as a covariate. P-values for the treatment-by-visit interaction were 0.7411 (overall), 0.5943 (doxepin 3 mg versus placebo), and 0.5612 (doxepin 6 mg versus placebo).

Table 4 presents the sponsor’s results for sTST when the average over the 2 nights of each visit is the dependent variable. The sponsor prespecified the sTST as a key secondary endpoint rather than the sWASO.

Table 4 Study 501: sTST averaged over 2 nights of each Visit

Timepoint	Doxepin 3 mg		Doxepin 6 mg	
	LS Mean Difference from Placebo (Std. Err)	p-value	LS Mean Difference from Placebo (Std. Err)	p-value
Overall	11.9 (5.97)	0.0469	17.3 (5.96)	0.0042
By Visit				
Visit 4 (Days 2 & 3)	14.7 (6.76)	0.0307	20.8 (6.79)	0.0025
Visit 5 (Days 16 & 17)	12.5 (7.86)	0.1132	16.5 (7.82)	0.0359
Visit 6 (Days 30 & 31)	8.6 (8.83)	0.3317	14.5 (8.82)	0.1023

Notes: Subjects provided sTST values on both days of the study visit; data from paired days were averaged prior to analysis. See SCE [Addendum Attachment 1 Section 6.1](#) for an overview of the extent of missing data by study visit. Values of sTST >8 hours are excluded from the analysis (no value met this criterion; see SCE [Addendum Attachment 1 Table A1](#) for details). P-values comparing each active treatment versus placebo are obtained from a MMRM analysis that includes the main effects for treatment, visit, and the treatment-by-visit interaction with the baseline value as a covariate. P-values for the treatment-by-visit interaction were 0.9609 (overall), 0.8151 (doxepin 3 mg versus placebo), and 0.8050 (doxepin 6 mg versus placebo).

Table 5 summarizes the sponsor’s post-hoc analyses using MMRM. It presents the treatment difference averaged over the entire treatment period since the treatment by time interaction is not statistically significant at the 0.10 level.

Table 5 Summary of Sponsor’s Results for Overall Treatment Period from Post-Hoc MMRM

Study	Measure	Doxepin Dose Group ¹	P-value from test of treatment-by-time interaction	Estimated Difference from Placebo	
				LS Mean Difference (Std.Err)	p-value
SP-0501	sWASO	All doses	0.7411		
		3 mg	0.5943	-10.2 (4.41)	0.0213
		6 mg	0.5612	-14.2 (4.41)	0.0014
SP-0501	sTST	All doses	0.9609		
		3 mg	0.8151	11.9 (5.97)	0.0469
		6 mg	0.8050	17.3 (5.96)	0.0042
SP-0503	sWASO	All doses	0.2816		
		3 mg	0.6254	-18.3 (6.49)	0.0052
SP-0503	sTST	All doses	0.2046		
		3 mg	0.5720	18.9 (7.41)	0.0114

1. “All doses” represents the overall test of the treatment-by-time interaction.

MMRM Analysis: SP-0501 (Using the Study Day as the Repeated Measurement)

Instead of taking only the first night for the analysis as preplanned or averaging over the 2 nights of each visit it is possible to perform an MMRM analysis using all 6 days of post-baseline data without averaging before performing the analysis.

Results from the MMRM analysis for sWASO from SP-0501 using the study day as the repeated measurement are provided in Table 6. For this analysis, each study day was used to define the time period for the repeated measurements. Though it is typical to average over paired study days prior to analysis, using the study day as the repeated measure is a reasonable sensitivity analysis to demonstrate the robustness of the model to the choice of time period for repeated measurements.

The treatment-by-time interaction was not significant (p=0.2047). Also not significant was the test that the difference between doxepin 3 mg and placebo was constant over time (p=0.1165) and the test that the difference between doxepin 6 mg and placebo was constant over time (p=0.1677). The tests for interaction meet the post-hoc criteria to support the conclusion that the effects of treatment do not differ significantly across time; however, the p-values associated with the two components of the treatment-by-time interaction are smaller than those obtained when averaging data from the two paired-days within each study visit prior to analysis. The smallest observed treatment effect for doxepin 6 mg was -3.4 minutes at Day 30; this was the only day for which the estimate for sWASO for doxepin 6 mg was not reduced by at least 10 minutes when compared to placebo. In fact, the estimated treatment effect on the following day was -23.1 minutes (Day 31). In the sponsor’s opinion, these data are consistent with the fact that sleep patterns in patients with insomnia have substantial night-to-night variability, and averaging data from the paired nights, although they prespecified not doing so, reduces the inherent variability seen in insomnia patients. Nonetheless, the maximum observed effect for doxepin 6 mg compared to placebo occurred on Day 31, the last day sWASO was obtained in this trial.

Table 6 Post-Hoc MMRM analysis of sWASO by Study Day

Timepoint	Doxepin 3 mg		Doxepin 6 mg	
	LS Mean Difference from Placebo (Std. Err)	p-value	LS Mean Difference from Placebo (Std. Err)	p-value
Overall	-10.0 (4.41)	0.0249	-14.2 (4.41)	0.0015
By Day				
Day 2	-19.9 (5.67)	0.0006	-19.4 (5.70)	0.0008
Day 3	-6.8 (6.09)	0.2676	-16.5 (6.11)	0.0073
Day 16	-11.3 (7.98)	0.1596	-10.2 (7.96)	0.2028
Day 17	-9.0 (6.72)	0.1831	-12.5 (6.73)	0.0647
Day 30	0.4 (7.36)	0.9559	-3.4 (7.34)	0.6443
Day 31	-13.3 (7.23)	0.0678	-23.1 (7.23)	0.0016

Notes: Subjects provided sWASO values at each day (see SCE [Addendum Attachment 1 Section 6.1](#) for an overview of the extent of missing data by study day). Values of sWASO >8 hours are excluded from the analysis (one value met this criterion; see SCE [Addendum Attachment 1 Table A1](#) for details). P-values comparing each active treatment versus placebo are obtained from a MMRM analysis that includes the main effects for treatment, day, and the treatment-by-day interaction with the baseline value as a covariate. P-values for the treatment-by-day interaction were 0.2047 (overall), 0.1165 (doxepin 3 mg versus placebo), and 0.1677 (doxepin 6 mg versus placebo).

The corresponding MMRM analysis of sTST by study day is presented in Table 7.

Table 7 Post-Hoc MMRM analysis of sTST by Study Day

Timepoint	Doxepin 3 mg		Doxepin 6 mg	
	LS Mean Difference from Placebo (Std. Err)	p-value	LS Mean Difference from Placebo (Std. Err)	p-value
Overall	11.4 (5.97)	0.0570	17.2 (5.95)	0.0042
By Day				
Day 2	21.6 (8.72)	0.0141	21.9 (8.76)	0.0131
Day 3	7.8 (7.87)	0.3249	19.6 (7.90)	0.0140
Day 16	14.4 (10.36)	0.1671	18.9 (10.29)	0.0674
Day 17	9.4 (8.30)	0.2577	13.9 (8.29)	0.0941
Day 30	2.3 (10.04)	0.8209	9.9 (9.99)	0.3241
Day 31	13.1 (10.61)	0.2172	19.0 (10.55)	0.0737

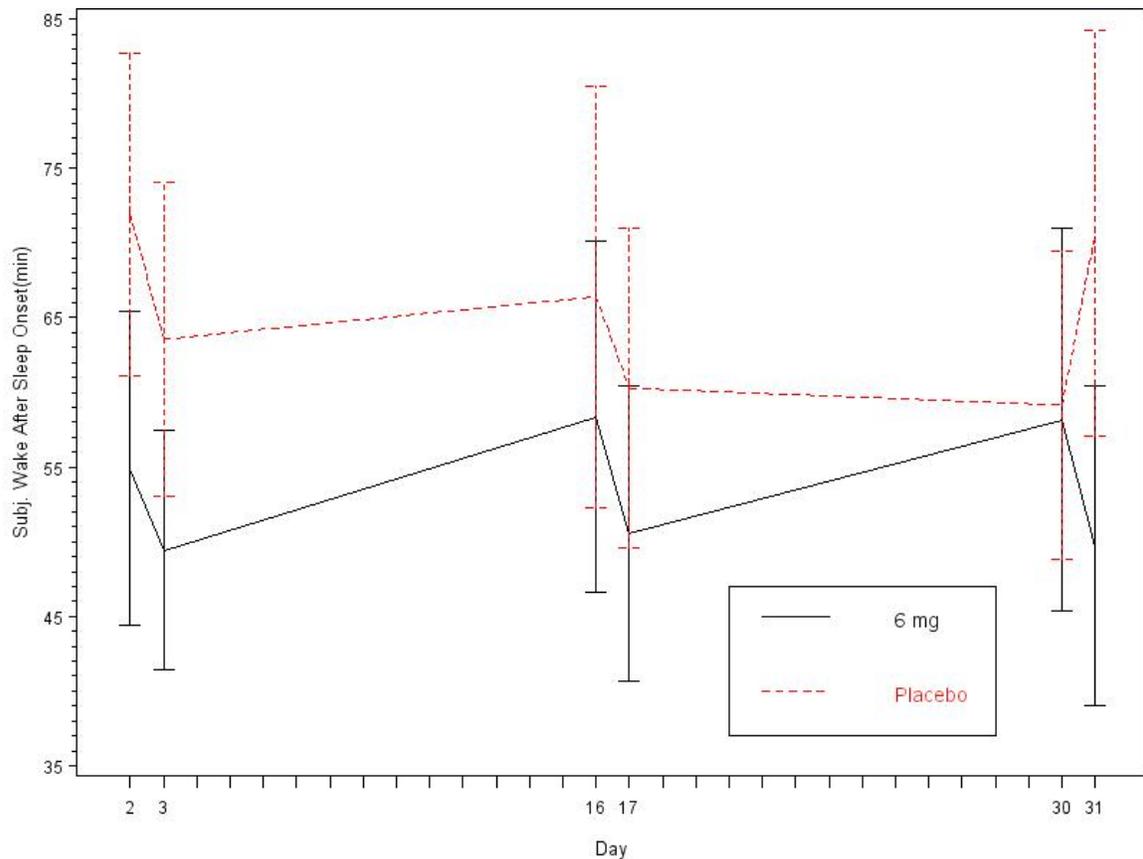
Notes: Subjects provided sTST values at each day (see SCE [Addendum Attachment 1 Section 6.1](#) for an overview of the extent of missing data by study day). Values of sTST >8 hours are excluded from the analysis (no value met this criterion; see SCE [Addendum Attachment 1 Table A1](#) for details). P-values comparing each active treatment versus placebo are obtained from a MMRM analysis that includes the main effects for treatment, day, and the treatment-by-day interaction with the baseline value as a covariate. P-values for the treatment-by-day interaction were 0.8970 (overall), 0.5607 (doxepin 3 mg versus placebo), and 0.8878 (doxepin 6 mg versus placebo).

4.2. Reviewer's Comments

Throughout this review this reviewer's focus is on the 6 mg vs. placebo comparisons since 3 mg vs. placebo differences were even less consistent over time.

Figure 1 shows the mean subjective WASO by day for the six days at which it was assessed. Subjects rated their WASO for the previous night so, for example, Day 2 corresponds to Night 1 and so on. Lines were drawn between the within group means at different days and vertical lines indicate the location of two standard errors from the group mean. It is noticeable that there is a jump between days 30 and 31 in both groups which is bigger than between previous consecutive days and in the case of this visit, unlike the others, both groups' jumps favor the drug. At day 30 the group means are very close, whereas, at day 31 they are far apart. A within treatment group test to see whether the mean sWASO varies by visit (or day) was nominally significant for the placebo group for the analysis based on the first night of each visit as well as the analysis based on all 6 days (0.0270 and 0.0393, respectively). Corresponding analyses using the 2 night average as endpoint were not nominally significant, nor were similar tests within the 6 mg group nominally significant regardless of the times used in the analysis. A test for a difference between just days 30 and 31 within the placebo group is nominally significant, $p=0.0365$. The day 30 6mg vs. placebo treatment difference and the day 31 6 mg vs. placebo are nominally significantly different, $p=0.0107$. In order to justify use of the 2-day average as the endpoint for analysis we would generally expect that the 2 nights have the same mean but the data suggests that they may not. Regardless of this, the first night of each visit was prespecified as the timepoint for the primary analysis, i.e., not the 2-day average.

Figure 1 Mean Subj. WASO over Time



Although the sponsor did not include center effects in their MMRM models this reviewer included them in MMRM analyses since they were included in the corresponding prespecified ANCOVA model and tests of significance for center effects were statistically significant. The p-value for the null hypothesis that all center effects are zero is 0.0066, which suggests that they are not. Note that it doesn't seem that there are major differences in the assessment of treatment effect (or treatment by time interaction) between the MMRM analyses adjusted for center and those not adjusted for center.

Note that days 2, 16, and 30 are the first days of each 2 day visit and the protocol specified that the analysis was to be based on only the first day of each visit. Neither the day 16 nor the day 30 treatment difference between 6mg and placebo is nominally significant ($p=0.188$ and 0.6119) based on a mixed model (MMRM) analysis of SWASO over days 2, 16, and 30. Note that these were the original times designated for analysis but the sponsor focused on the average of paired nights in their most recent submission. If we average the differences between 6 mg and placebo over days 16 and 30 that estimated difference, 7.1 with a standard error of 6.1, is not nominally significant, $p=0.2456$. If we average over day 2 as well, then we get the average difference between 6mg and placebo over the whole double blind period (first night of each visit). This estimated difference, 11.3 with a standard error of 4.96, does reach the nominal level, $p=0.0243$.

Based on an MMRM analysis of the first night (the time selected for analysis in the protocol) of each visit the estimated treatment difference between 6 mg and placebo at day 30 is 3.61 (S.E.=7.11), $p=0.6119$. This model included baseline SWASO as a covariate, as well as center effects, treatment, visit and treatment by visit interaction effects. The covariance structure for repeated measurements within the same patient was specified as unstructured (no presumed time pattern for the correlations and variances). The MMRM analysis focused on the second night of each visit found a treatment difference of 22.9 (S.E.=7.13) at day 31, $p=0.0016$. Another possible way of handling the two nights of each visit is to average the response over the two nights and use that average as the response (dependent) variable in the analysis, i.e., the response for a particular visit is the average of the two nights associated with that visit. Although, as noted above, one would typically expect that the two nights have the same mean which it appears may not be the case here. Based on an MMRM analysis of this two night average data the estimate for the treatment difference at the last visit is 13.46 (S.E.=6.08) which has a p-value of 0.0278. This difference based on the average of two nights would not be significant if a Bonferroni adjustment had been specified to permit analyzing the data from the visits in different ways (e.g., $\alpha=0.025$ significance level for looking at first night and average of two nights or alternatively, $\alpha=0.0167$ for looking at first night, second night, and average of 2 nights). Had a Bonferroni adjustment for looking at the three different summaries of the data from each visit been prespecified the day 31 result (based on the MMRM model for the second night of each visit) would be statistically significant as a result of the 2nd night analysis. However, the difference is not significant at Visit 5 after Bonferroni adjustment. This would impact our conclusions for Visit 6 if we tested the treatment difference at each visit in order of visits starting with Visit 4.

Table 8 Study 501: Unadjusted P-values for Analyses of SWASO (6 mg vs. Placebo)

Analysis /Missing Data Handling	Analysis Variable for SWASO	Visit 5 (Night 15, 16)	Visit 6 (Night 28, 29)
MMRM OC	1 st Night	0.1743	0.6119
	2 nd Night	0.0601	0.0016
	Average of 2 Nights	0.0443	0.0278
ANCOVA OC	1 st Night	0.2017	0.6282
	2 nd Night	0.0418	0.0010
	Average of 2 Nights	0.0419	0.0199
ANCOVA w/ LOCF	1 st Night	0.1966	0.8690
	2 nd Night	0.0532	0.0008
	Average of 2 Nights	0.0460	0.0221

In addition, there may also be a multiplicity issue associated with the choice of “primary” subjective maintenance endpoint since the sponsor prespecified sTST over sWASO but the minutes of the pre-NDA meeting suggest that the Division seemed to have voiced a preference for sWASO.

None of the Visit 6 comparisons of 6 mg to placebo are nominally significant for the sTST except for the LOCF ANCOVA analysis of the 2nd night only sTST. However, the first night of each visit was prespecified as the data to be used for the analysis. If the Bonferroni method had been prespecified to adjust for the multiplicity of looking at different combinations of handling the multiple nights of each visit then this p-value would not be statistically significant either. Thus, it seems that the evidence for a long term effect on sTST is weak and note that the sponsor prespecified the sTST as the primary measure of subjective maintenance rather than the sWASO.

Table 9 Study 501: Unadjusted P-values for Analyses of sTST (6 mg vs. Placebo)

Analysis /Missing Data Handling	Analysis Variable for STST	Visit 5 (Night 15, 16)	Visit 6 (Night 28, 29)
MMRM OC	1 st Night	0.0772	0.3847
	2 nd Night	0.1094	0.0681
	Average of 2 Nights	0.0389	0.1065
ANCOVA OC	1 st Night	0.0870	0.4706
	2 nd Night	0.1183	0.1131
	Average of 2 Nights	0.0472	0.1464
ANCOVA w/LOCF	1 st Night	0.0554	0.2523
	2 nd Night	0.0754	0.0380
	Average of 2 Nights	0.0428	0.0783

The following analyses of the difference averaged over several visits were not pre-planned. They are just provided for comparison with the sponsor’s post-hoc analyses of the difference averaged over the entire double blind period since they may be more relevant to the difference at the end of double blind treatment. The recommended approach for assessing this is not averaging over visits, i.e., instead just using the data from the last visit. The 6mg vs. placebo difference averaged over the first night of visit 5 and visit 6 using MMRM is not significant for sWASO: 7.18 (5.87), p=0.223 or for sTST: 13.54 (8.66), p= 0.1194. Averaged over the second nights it reaches the nominal significance level for sWASO: 17.6 (5.76), p=0.003 and sTST: 16.03 (7.58), p=0.0358. Based on the two night average for each visit the treatment difference on sWASO reaches

nominal significance: 12.49 (4.93) $p=0.0122$; as does the difference on sTST: 15.16 (7.17), $p=0.0358$.

Assessment of whether the Treatment Difference varies with Time (Interaction)

It is common to use a test of interaction between two effects, such as treatment and time, based on a model including the two effects as well as effects for their interaction to check for support of a reduced model with the same effects except for the interaction. However, it is not strictly correct to conclude that a non-significant test for interaction implies a lack of any interaction. This is because a basic concept of hypothesis testing is that failure to reject the null hypothesis does not necessarily imply that the null hypothesis is true; rather, one can only say that there is not enough evidence to reject the null. Therefore, just because the test for interaction between time and treatment did not suggest significance at the 0.05 (or 0.010) level it does not allow us to conclude that there is zero interaction. A noninferiority approach or, actually, an equivalence approach to the interaction test (because we need to rule out a difference in either direction) seems more appropriate for the question of interest, namely whether or not the interaction is small enough to be unimportant. This is because this approach controls the probability of falsely rejecting the null hypothesis that there is interaction. This is straightforward if we only consider the treatment differences between two timepoints but becomes more complicated for more than two timepoints. This is because for more than two timepoints the margin may depend on some combination of the differences we want to allow at each of several timepoints. Of course, such an equivalence approach really needs to be planned in advance with an equivalence tolerance margin but we can informally examine whether the interaction effects exceed a few reasonable levels. Before we look at equivalence testing for the interaction we will look at the standard interaction test where the null hypothesis is that the interaction effects are all zero.

This reviewer found that the p-value for the difference between the day 2 and day 30 treatment differences is 0.0528 (based on only 6mg and placebo data). Adjusting for center and including all three groups the same difference has a p-value of 0.0487. The estimated difference between the treatment differences at the two times is 16.30 with a 95% confidence interval of (0.10, 32.50). The 90% confidence interval is 2.72 to 29.87. That means that we couldn't rule out a difference of nearly 30 minutes. This would seem to exceed any reasonable equivalence margin. Also, if visit is entered into the model as continuous, i.e., we assume the subjective WASO is a linear function of time, then the test for interaction between treatment and visit is $p=0.0524$ based on the data from days 2, 16, and 30. It seems inappropriate to make an inference about the effect at a particular time by averaging over all times unless one knows for sure that the effect is the same at all times (and times between them).

The p-value for the standard test of interaction between treatment and time (first night of each visit) on SWASO excluding the 3mg group is 0.1414. If we include all days for each visit, for a total of 6 days, the p-value for the test of interaction between treatment and time on SWASO is 0.2692. However, this reviewer found that the simulated power to detect an interaction at a significance level of 0.05 between treatment and time, assuming the observed differences in treatment effect between days 2, 16, and 30 on the subjective

WASO are true, is just 43%. This power assuming the observed model is true is sometimes called the post-hoc power. For a significance level of 0.10 it is 55% in this case. Therefore, even if one could conclude that the null hypothesis were true if the alternative was not rejected, the test for interaction is likely underpowered so it may only appear that the null hypothesis is true because of the high variability associated with the small sample size. Based on simulations in order to get 90% power for the interaction test at the 0.05 significance level with the given sample size and the observed effects we would need to reject the hypothesis of no interaction whenever $p \leq 0.54$. To get 80% power we would need to reject whenever $p \leq 0.33$. In terms of the size of the effect needed to have 80% power to detect the interaction at the 0.05 significance level we would need, for example, a difference of 26.2 minutes between the treatment group differences at visit 5 and visit 6 and the treatment group differences equal at visit 4 and visit 6 (which also implies a difference of 26.2 between visits 4 and 5). The differences needed are not unique but follow an equation (see footnote 1) that is quadratic in each of the two treatment by time interaction parameters: one for the difference in the treatment effects between visits 4 and 6 and one for the difference between visits 5 and 6. The estimated number of patients needed to detect an interaction the size of the estimated interaction effects (-16.1 and -7.1) with 80% power at a 0.05 significance level is 170 per group which is a little more than twice the size of study 501. At the 0.10 significance level for the interaction test 125 patients would be needed per group which is still 60% more than the size of the study. For the two-night-average endpoint data the variability is a little smaller so the effects necessary for 80% power are a little smaller, for example, a difference of 17.1 between the treatment group differences at visit 5 and visit 6 and the treatment group differences equal at visit 4 and visit 6. The post-hoc power for the interaction would be based on the estimated interaction parameters: -4.5 and 1.9 which are much smaller than 17.1. Therefore, the post-hoc power of 16% for a 0.05 significance level and 31% for a 0.10 significance level is much less than 80%. Clearly, this test of the interaction is evaluated post-hoc and the study was not adequately powered to give a fair assessment of its significance.

Another remaining question is did the design have enough coverage of the 30 day range to detect a non-constant relationship, i.e., we can't necessarily assume that the treatment difference is constant at all times in between scheduled assessment times. If it is not then it wouldn't make much sense to average over the three assessments to estimate the effect at the end, whether or not the effect is constant at these three times.

Equivalence Testing for the Treatment by Time Interaction Effects

The F test statistic that the sponsor used to evaluate the null hypothesis that all interaction terms are zero can be used for an equivalence test as well. Under a non-equivalence null hypothesis at least one of the interaction terms is greater than, e.g., 10 minutes and the F statistic has a non-central F distribution. We can determine the non-centrality parameter from the theory for mixed models¹. We can implement this in the current situation if we assume that the observed parameter estimates are true for all parameter estimates other

¹ The noncentrality parameter takes the form $\beta^T k (k^T [X^T R^{-1} X]^{-1} k)^{-1} k^T \beta$ where k is the contrast associated with the F test for interaction, β is the parameter vector for the fixed model, X is the design matrix, and R is the covariance matrix for the dependent variable.

than the interaction effects and we assume the interaction effect parameters are those associated with the null hypothesis (closest to the alternative hypothesis). If we then compare the observed F statistic to the non-central F distribution, which holds under the null hypothesis, we can perform the equivalence test. In particular, if the interaction terms are all zero then the observed F statistic will tend to be smaller than what would be observed if the interaction terms are not all zero. The idea is most easily illustrated for an ANOVA model including terms for treatment, visit, and treatment by visit interaction effects. In this case under the null hypothesis the expected value of the numerator of the F statistic is the sum of two terms, a term based on the error variance and a term proportional to the sum of the squared interaction parameters. The expected value of the denominator is the same as the first term in the numerator. Therefore, because the second term is always non-negative the F statistic will tend to be bigger when the null hypothesis of non-equivalence is true than when it is not. The F statistic has this same property under the more complicated mixed model for repeated measures. So, if the observed F is less than the 0.05 percentile of the noncentral F distribution then we can reject the null hypothesis that at least one interaction term is greater than 10. Applying this method to the study 501 subjective WASO data from the first night of each visit yields a p-value of 0.6092, i.e., we can not reject the null hypothesis that at least one of the interaction terms is greater than or equal to 10. A similar result was also obtained using a different test based on the maximum of two correlated MMRM model based t-statistics: the first associated with the difference between the treatment effects at days 2 and 16, and the other associated with the difference between treatment effects at days 2 and 30. A difference of 10 minutes seems important considering that it is about 50% of the estimated treatment difference observed for Night 1 (Day 2). The noncentral F statistic was used to evaluate the Treatment by Time Interaction effects for various analysis timepoints. The results are shown in Table 10.

Table 10 Study 501: Equivalence Test for Treatment by Time Interaction Effects

Analysis Timepoints for SWASO	Margin	p-value*
First Night	10	0.6092
	15	0.3663
	20	0.1617
Second Night	10	0.2502
	15	0.0797
	20	0.0147
Average over 2 nights	10	0.1073
	15	0.0157
	20	0.0010
All 6 Days	10	0.2684
	15	0.0676
	20	0.0080

*If p for non-central F statistic <0.05 then can conclude interaction effects are smaller than margin otherwise cannot

5. Study 503

5.1. Subjective WASO and Subjective TST

Study 503 included Doxepin 1 and 3 mg groups, as well as placebo, and it was conducted in the elderly.

The sponsor states that the test for interaction between treatment and visit on the sWASO yielded a p-value of 0.6254 (3mg vs. placebo) so they claim that the treatment effect does not vary significantly across visits and they present an overall average treatment difference. However, failing to reject the null hypothesis that all interaction terms are zero may just mean there is insufficient evidence to reject the null due to a lack of power. In order to claim that all interaction effects are small enough to be unimportant we would need to use an equivalence testing approach on the interaction terms. If we use the null hypothesis that at least one of them is as large as 10 minutes, then, using the same method as used above for study 501, we get $p=0.2147$. If we use the null hypothesis that at least one of them is as large as 15 then $p=0.1124$. Thus, it seems that contrary to the sponsor's conclusion of no interaction the visit specific treatment effects may vary by as much as 15 minutes. Therefore, we need to include the treatment by visit interaction in the model in order to reliably estimate the treatment effect at each visit.

Based on a mixed model for repeated measures the treatment difference between 3 mg and placebo on sWASO was not nominally significant at day 58 (8.7, $p=0.3111$), day 30 (19.0, $p=0.0512$), or day 16 (16.8, $p=0.0784$) but was at day 86 (24.1, $p=0.0141$) and day 2 (21.2, $p=0.0269$). This model included effects for baseline sWASO, center, treatment, visit, and treatment by visit interaction and assumed a general structure for the within patient covariance of the repeated measurements. The 3mg placebo difference averaged over the last two visits had a p-value of 0.0503 (15.9 [S.E.=8.1]). This did reach nominal significance if averaged over the last 3 or 4 visits ($p=0.0201$ and 0.0136, respectively). These analyses of the difference averaged over several visits were not pre-planned. They are just provided for comparison with the sponsor's post-hoc analyses of the difference averaged over the entire double blind period ($p\leq 0.01$ [Table 1]) since they may be more relevant to the difference at the end of double blind treatment. The recommended approach for assessing this is not averaging over visits, i.e., instead just using the data from the last visit. The only difference between the results based on ANCOVA and MMRM is in the nominal significance of the treatment difference at day 2. The prespecified ANCOVA model found an insignificant treatment difference there (19.2 [S.E.=10.0], $p=0.0561$ OC). Under a closed testing procedure multiplicity adjustment for testing multiple times testing would have to stop after day 2 for the MMRM analyses because of the insignificant result at day 16.

The MMRM results for sTST were $p=0.0329$ for day 86, $p=0.3618$ for day 58, $p=0.0397$ for day 30, $p=0.3558$ for day 16, and 0.0457 for day 2. The 3mg placebo difference averaged over the last two visits had a p-value of 0.0750 (16.2 [S.E.=7.6]). This did reach nominal significance if averaged over the last 3 or 4 visits ($p=0.0295$ and 0.0332, respectively). The sponsor's result for 3mg vs. placebo over the entire treatment period was $p\leq 0.05$ (Table 2). For the original ANCOVA OC analyses of sTST the 3 mg vs.

placebo difference on the subjective total sleep time, which the sponsor prefers over subjective WASO and which they specified as key secondary, was not significant at day 86 (19.7 [S.E.= 11.0] , p=0.0752), day 58, 16, or day 2 (19.0 [S.E.=11.0], p=0.0865). Based on LOCF imputation the ANCOVA analysis of 3 mg vs. placebo at day 86 yielded an estimated difference of 22.5 (10.4 S.E.), p=0.0310 but in Insomnia where first night drug effects are usually expected LOCF can be misleading about the duration of effect. Even if one accepts the LOCF result for day 86, there is still the inconsistency of day 58 and day 16. Therefore, study 503 doesn't provide much evidence of a persistent statistically significant effect of 3mg Doxepin as compared to placebo on subjective maintenance.

Sponsor's Assessment of IVRS subjective Total Sleep Time data

A separate, more extensive assessment of sTST was done by an interactive voice response system but this was not implemented until the study was already underway. The self-rated IVRS for sTST in Study SP-0503 (elderly study) provided additional support for subjective sleep maintenance. Because IVRS data were not available for all subjects (Statistical Methods – Phase 3 Chronic Insomnia Studies), analyses were conducted two ways: using observed data only and imputing missing baseline values using the overall population mean at baseline. Analyses using observed data only showed statistically significant increases in mean sTST for doxepin 3 mg versus placebo at all weeks other than Week 2 (p=0.0590), Week 7 (p=0.0513), and Week 9 (p=0.3723). For analyses where missing baseline values were imputed (n=31 subjects) using the overall mean value at baseline, there were statistically significant increases in mean sTST for doxepin 3 mg versus placebo at all weeks other than Week 9 (p=0.4525). Notably, although not all subjects in the study had data available from the IVRS (49 subjects were randomized prior to the IVRS activation; an additional 31 subjects who were activated into the IVRS are missing baseline values), the two analyses are consistent with each other and the treatment effect is consistent for doxepin 3 mg versus placebo across the entire duration of the study with the exception of one week (Week 9) where the p-value comparing doxepin versus placebo is 0.37 for the observed data and 0.46 for the imputed data.

Reviewer's Comments on IVRS subjective TST data

A significant imbalance in the efficacy measure at baseline (p=0.0314 here) is a serious issue which may undermine any post-baseline comparisons. If the groups were not equivalent at baseline then any observed post-baseline treatment group difference may be due to the baseline imbalance rather than the treatment. The 3 mg group had a higher mean STST than placebo at baseline: 313 vs. 276 (medians: 330 vs. 300). Ignoring that for the moment, observed data still did not demonstrate a persistent nominally significant treatment difference. The difference between 3mg and placebo was nominally significant at weeks 1, 3, 4, 5, 6, 8, 10, 11, and 12 but was not at week 2, 7, or 9. In another analysis the sponsor imputed missing baseline scores with the overall mean baseline score. This tends to artificially make the groups more alike even though on the basis of the observed baseline scores they were significantly different. Eighty two subjects were randomized to 3 mg and 81 were randomized to placebo. Forty nine subjects (16 from 3 mg, 14 from 1 mg, and 19 placebo) have missing data because they were randomized prior to the IVRS

activation. Thirty one additional randomized subjects (11 from 3mg, 10 from 1 mg, and 10 placebo) had a missing baseline IVRS sTST score even though they were randomized after IVRS activation. Thus, 34% of randomized patients have no baseline IVRS STS score available. The observed case analysis is not strictly a randomized comparison, nor is the post-imputed baseline analysis because there are still 49 pre-IVRS randomized patients excluded. The overall mean baseline score may be different if the baseline score were known for all randomized subjects so the result based on the imputed baseline would change as well. The quality of this IVRS data seems to be substandard and it is on an endpoint that was not designated as a key secondary or even a secondary so it is difficult to assign multiplicity adjusted p-values to these analyses. Therefore, it seems difficult to consider these IVRS data as providing much in the way of support of a subjective maintenance claim. Also, because this was a study in the elderly it does not directly address subjective maintenance for non-elderly adults where the evidence is most lacking.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22036	ORIG-1	SOMAXON PHARMACEUTICA LS INC	SILENOR (DOXEPIN HCL)

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
10/28/2009

KUN JIN
10/28/2009

HSIEN MING J J HUNG
10/29/2009



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: 22,036

Drug Name: Silenor (Doxepin HCl)

Indication(s): Insomnia

Applicant: Somaxon Pharmaceuticals

Date(s): Submission Date: January 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 1

Medical Division: Neurology (HFD-120)

Clinical Team: June Cai, M.D.
Devanand Jillapalli, M.D., Team Leader
Russell Katz, M.D., Division Director

Project Manager: Cathleen Michaloski

Keyword(s): Multiplicity

Table of Contents

LIST OF TABLES.....	4
LIST OF FIGURES.....	5
1 EXECUTIVE SUMMARY	6
1.1 CONCLUSIONS AND RECOMMENDATIONS	6
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	6
1.3 STATISTICAL ISSUES AND FINDINGS	7
2 INTRODUCTION	9
2.1 OVERVIEW.....	9
2.2 DATA SOURCES	10
3 STATISTICAL EVALUATION	11
3.1 EVALUATION OF EFFICACY	11
3.1.1 <i>Study 401</i>	11
3.1.1.1 Sponsor's Results.....	12
3.1.1.2 Reviewer's Comments	14
3.1.2 <i>Study 402</i>	16
3.1.2.1 Patient Disposition and Baseline Demographics.....	16
3.1.2.2 Sponsor's Results.....	16
3.1.3 <i>Study 501</i>	19
3.1.3.1 Patient Disposition	23
3.1.3.2 Baseline Demographic and Disease Characteristics.....	24
3.1.3.3 Sponsor's Results.....	25
3.1.3.4 Reviewer's Results.....	31
3.1.4 <i>Study 502</i>	38
3.1.4.1 Sponsor's Results.....	39
3.1.4.2 Reviewer's Comments	43
3.1.5 <i>Study 503</i>	43
3.1.5.1 Study Design and Statistical Methods.....	43
3.1.5.2 Patient Disposition	45
3.1.5.3 <i>Sponsor's Results</i>	47
3.1.5.4 Reviewer's Comments	51
3.1.6 <i>Study 509</i>	56
3.1.6.1 Study Design and Statistical Methods.....	56
3.1.6.2 Patient Disposition	57
3.1.6.3 Baseline Demographics.....	58
3.1.6.4 Sponsor's Results.....	59
3.1.6.5 Reviewer's Results.....	62
3.2 EVALUATION OF SAFETY	63
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	63
4.1 GENDER, RACE AND AGE	63
4.1.1 <i>Gender</i>	64
4.1.2 <i>Race</i>	65
4.1.3 <i>Age</i>	66
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	67
4.2.1 <i>Treatment Group Differences by Site</i>	67
5 SUMMARY AND CONCLUSIONS	71
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	71

5.2 CONCLUSIONS AND RECOMMENDATIONS73

LIST OF TABLES

Table 1 Overview of Clinical Studies	7
Table 2 Summary of Key Analysis p-values by Study	8
Table 3 Study 401: Analyses of Wake Time During Sleep	13
Table 4 Study 401: Analyses of Wake Time After Sleep Onset.....	13
Table 5 Study 401: Analyses of Latency to Persistent Sleep.....	14
Table 6 Study 402: Analyses of Wake Time During Sleep	17
Table 7 Study 402: Analyses of Wake Time After Sleep Onset.....	17
Table 8 Study 402: Analyses of Latency to Persistent Sleep.....	18
Table 9 Study 501 Patient Disposition	23
Table 10 Study 501: Demographic and Other Baseline Characteristics: Safety Analysis Set.....	24
Table 11 Study 501: Primary Efficacy Variable – WASO on Night 1: ITT Analysis Set.....	25
Table 12 Study 501: WASO on Night 15, Night 29, and the Average of Nights 1, 15, and 29:	26
Table 13 Study 501: LPS at Baseline, Night 1, Night 15, Night 29, and Average of Nights 1, 15, and 29: ITT Analysis Set.....	27
Table 14 Study 501: Subjective Sleep Variables sTST, sWASO, and sNAASO at Baseline, Day 2, Day 16, and Day 30: ITT Analysis Set.....	28
Table 15 Study 501: Subjective Sleep Variables LSO and Sleep Quality at Baseline, Day 2, Day 16, and Day 30: ITT Analysis Set.....	29
Table 16 Study 501: SE All Night Over Time: ITT Analysis Set	30
Table 17 Comparison of OC and LOCF analyses for objective WASO.....	32
Table 18 Analyses of subjective WASO by Night (Observed Cases Analyses).....	33
Table 19 Study 501: Subjective WASO: Comparison of OC and ITT-LOCF analyses	34
Table 20 Study 501: Analyses of objective LPS by Night (Observed Cases Analyses).....	35
Table 21 Study 501: Wake Time After Sleep by Night (Observed Cases).....	36
Table 22 Study 501: Results for 1 st Night, 2 nd Night, and Average of two at each Visit (OC).....	37
Table 23 Study 502: Baseline Demographics.....	40
Table 24 Study 502: Primary Efficacy Variable – LPS on Night 1: ITT Analysis Set.....	40
Table 25 Study 502: WASO on Night 1: ITT Analysis Set.....	41
Table 26 Study 502: SE Overall and by First, Second, and Final Third of the Night on Night 1: ITT Analysis Set ..	42
Table 27 Study 503: Patient Disposition	46
Table 28 Study 503: Baseline Demographics.....	47
Table 29 Study 503: Objective WASO on Night 1 in ITT population	48
Table 30 Study 503: Subjective TST at Baseline, Day 2: ITT Analysis Set.....	49
Table 31 Study 503: LPS over Time in ITT population (Sponsor’s Results).....	50
Table 32 Study 503: SE in Hour 8 at Baseline, Night 1: ITT Analysis Set	51
Table 33 Study 503: Number of Awakenings after Sleep Onset in ITT population	51
Table 34 Study 503: Comparison of Results from various Imputation Methods for Missing objective WASO data ..	52
Table 35 Study 503: Objective Latency to Persistent Sleep by Night for OC and ITT-LOCF analyses	53
Table 36 Study 503: Comparison of Observed Case and ITT-LOCF results for subjective WASO by Night.....	54
Table 37 Study 503: Wake Time After Sleep by Night (OC).....	55
Table 38 Study 509: Patient Disposition, All Randomized Patients.....	58
Table 39 Study 509: Demographic and other baseline characteristics, Safety Set	59
Table 40 Study 509: Primary Efficacy Variable – Subjective Total Sleep Time – at Baseline and Week 1: ITT Analysis Set.....	60
Table 41 Study 509: Latency to Sleep Onset by Week: ITT Analysis Set (OC)	61
Table 42 Study 509: Subjective WASO analyses for OC and ITT-LOCF populations by Week.....	62
Table 43 Study 509: Subjective LSO analyses for OC and ITT-LOCF populations by Week.....	63
Table 44 Study 501: Gender Subgroup Analyses of Objective WASO on Night 1.....	64
Table 45 Study 503: Gender Subgroup Analyses of Objective WASO on Night 1.....	64
Table 46 Study 509: Gender Subgroup Analyses of Subjective WASO over Week 1	64
Table 47 Study 502: Gender Subgroup Analyses of Objective LPS on Night 1	65
Table 48 Study 501: Race Subgroup Analyses of Objective WASO on Night 1	65

Table 49 Study 503: Race Subgroup Analyses of Objective WASO on Night 1	65
Table 50 Study 509: Race Subgroup Analyses of Subjective WASO over Week 1	66
Table 51 Study 502: Race Subgroup Analyses of Objective LPS on Night 1	66
Table 52 Summary of Key Analyses by Study	72

LIST OF FIGURES

Figure 1 Study 501: SE by Hour of the Night on Night 1: ITT Analysis Set	31
Figure 2 Study 503: Objective WASO over Time	48
Figure 3 Study 501: 6 mg and Placebo Estimated Night 1 Difference in Objective WASO by Site	68
Figure 4 Study 502: 6 mg and Placebo Estimated Night 1 Difference in Objective Log (LPS) by Site	69
Figure 5 Study 503: 3 mg and Placebo Estimated Night 1 Difference in Objective WASO by Site	70
Figure 6 Study 509: 3 mg and Placebo Estimated Week 1 Difference in Subjective WASO by Site	71

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The clinical efficacy studies in this application seem to support an effect of the drug on sleep maintenance as measured by the Wake Time after Sleep Onset. However, there were no consistent effects of the drug on Latency to Persistent Sleep; no statistically significant differences between Doxepin and placebo in terms of LPS were observed beyond night 1 in any study. (b) (4)



1.2 Brief Overview of Clinical Studies

All investigators in the sponsor-identified key efficacy studies were based in the United States. Studies 401 and 402 were phase 2, 4 period, 4 treatment crossover studies. Efficacy measures were recorded during 2 nights in each period. For analysis purposes, the measures were averaged over the 2 nights in each period. The phase 3 studies: 501, 502, 503, and 509 were double blind, randomized, multi-center, placebo controlled, parallel group studies. Studies 501, 503, and 509 are the key efficacy studies for the chronic insomnia population. Study 501 compared 6 mg and 3 mg doses to placebo in non-elderly adults. Objective and Subjective efficacy measures were recorded for nights 1, 2, 15, 16, 29, and 30. For the analysis the sponsor chose to focus on nights 1, 15, and 29, i.e., the first night at each of the three visits rather than the second night or the average of the first and second nights at each visit. Study 503 compared 3 mg and 1 mg doses to placebo in an elderly population. Objective and Subjective efficacy measures were recorded on nights 1, 15, 29, 57, and 85. Study 509 compared 6 mg and placebo in an elderly population. Subjective efficacy measures were recorded nightly for 28 days. For analysis purposes weeks 1 through 4 were examined by averaging measures within the same week over the week for each patient. Some key characteristics of the studies are presented in Table 1.

Table 1 Overview of Clinical Studies

Study	Groups (N)	Primary Endpoint	Dropout %	Duration
401	4 period Crossover N=67 P,1,3,6 mg	WTDS, night 1,2 avg	1.4	4 periods 2 days on, 5-12 off trt
402(Elderly)	4 period Crossover N=76 P,1,3,6 mg	WTDS, night 1,2 avg	4	4 periods 2 days on, 5-12 off trt
501	P (76), 3mg (77), 6mg(76)	WASO, night 1	12, 12, 11	35 nights
502 (Transient Insomnia)	P (282), 6mg(283)	LPS, night 1	0,0	1 night, 1 day
503 (Elderly)	P(81), 1mg (77), 3mg(82)	WASO, Night 1	14, 9, 10	85 nights
509 (Elderly/Subjective measures only)	P(125), 6mg(130)	Subj TST, week 1	10, 5	28

WTDS=Wake time during sleep; WASO=Wake time after sleep onset; LPS=Latency to Persistent Sleep; TST=Total Sleep Time

1.3 Statistical Issues and Findings

The sponsor's prespecified analysis plans for dealing with type I error issues associated with multiple endpoints and multiple timepoints often differed from the approach preferred by the FDA Neurology division and recommended to the sponsor in the pre-NDA meeting. The division recommended testing the objective WASO, followed by the subjective WASO, the objective LPS, and finally the subjective LSO. In addition for each efficacy measure they recommended starting at the latest time and working backwards to the first time until an insignificant result was observed. When there were multiple doses in a study this procedure would start with the high dose first. The results of this procedure can be determined from Table 2. In study 501 the 6 mg group was significantly improved compared to placebo in terms of the objective WASO at each visit (1st of two nights). However, the 6 mg group was not statistically significantly better than the placebo group in terms of the subjective WASO at night 29 (1st night of the last visit). Therefore, in order to control the experiment wise type I error at 0.05 level testing must stop with this test, i.e., the results on the LPS and sLSO endpoints, as well as the results for 1mg vs. placebo comparisons, can only be considered exploratory. This reviewer found that if the sponsor had chosen to perform the comparison on the 2nd night of each visit or the average of nights 1 and 2 instead of on the first night only then the 6 mg group would have won at each time in terms of the subjective WASO. However, the choice of which night's data to base the hypothesis on introduces another layer of multiplicity if we consider something different than the sponsor prespecified. Even if they had chosen differently we see that they would not have won on the LPS because the 6mg vs. placebo difference was not significant at night 29. In fact, there is no evidence from any of the studies that there is a statistically significant effect on latency to persistent sleep beyond night 1.

Table 2 Summary of Key Analysis p-values by Study

Study	Endpoint	Dose Group	P-Values as Compared to Placebo					
			Night 85	Night 57	Night 29	Night 15	Night 1	
401(Phase 2 Crossover)	WASO	6					<0.0001	
	LPS	6					0.0397	
402 (Phase 2 Crossover/ Elderly)	WASO	6					<0.0001	
	LPS	6					0.1063	
	WASO	6			0.0007	0.0011	<0.0001	
		3			0.0173	0.0025	<0.0001	
	sWASO	6			0.6282	0.2016	0.0004	
		3			0.6483	0.1512	0.0003	
	LPS	6			0.8643	0.5921	0.0009	
		3			0.7995*	0.2271*	0.0058	
	sLSO	6			0.6511*	0.1451	0.0492	
		3			0.2365*	0.9071*	0.1259	
	502 (Transient Insomnia)	WASO	6					<.0001
		sWASO	6					0.0063
		LPS	6					<0.0001
		LSO	6					<0.0001
503 (Elderly)	WASO	3	<.0001	0.0029	0.0005	0.0069	<0.0001	
		1	0.0330	0.1662	0.0878	0.1945	0.0053	
	sWASO	3	0.0153	0.5627	0.0296	0.0729	0.0561	
		1	0.0037	0.7417	0.0531	0.8571	0.8497	
	LPS	3	0.0286*	0.0522*	0.5422*	0.8388	0.1079	
		1	0.6493	0.1870*	0.1268*	0.8046	0.5733*	
	sLSO	3	0.8479	0.9931	0.6544*	0.916	0.0860	
		1	0.2826	0.9631*	0.1798*	0.3567*	0.2304*	
509 (Elderly/ Subjective Only)	sWASO	6		0.0026 (Week 4)	0.0016 (Week 3)	0.0145 (Week 2)	<0.0001 (Week 1)	
	sLSO	6		0.6629 (Week 4)	0.4635 (Week 3)	0.4884* (Week 2)	0.1547 (Week 1)	

Note: The empty cells reflect the different lengths of study. In studies 401 and 402 the night 1 results are actually the results for the average of nights 1 and 2 as pre-specified by the sponsor in the analysis plan. For study 509 which analyzed weeks instead of individual nights the time corresponding to the analysis is displayed in the cell.

* numerically favors placebo

For study 501, the sponsor did not consider nights after the first night to be key hypotheses, i.e., to include them in the set of hypotheses which would have the chance of a single type I error, over all hypotheses in it, protected at 0.05. The sponsor named key secondary endpoints but did not state unambiguously that they were to be tested in order until an insignificant result was obtained.

For study 503 the sponsor specified a clear hierarchy for testing. The subjective Total Sleep Time (sTST) over night 1 was the next endpoint after the objective WASO had been tested at

each Visit. Because the comparison between 3 mg and placebo on the sTST over Night 1 was not significant ($p=0.0865$) no claims can be made on the lower endpoints in the hierarchy (LPS and Sleep Efficiency). Note that the 3mg vs. placebo comparison of LPS on night 1 was not significant ($p=0.1079$) either, which is even more reason that no claims of an effect on Sleep Efficiency are possible.

In summary, study 501 in non-elderly adults provides some evidence for the superior efficacy of the 6 mg dose compared to placebo for sleep maintenance as measured by objective WASO. The differences between 6 mg and placebo in terms of subjective WASO were not consistently significant. They were not significant at the sponsor's prespecified key timepoints, first night of each visit, but they were nominally significant at the second night of each visit as well as for the average of the two nights at each visit. Although, the 3 mg dose was also nominally significantly better compared to placebo for sleep maintenance as measured by objective WASO the multiplicity adjustment requires us to consider these results as exploratory because of the insignificant LPS results for the 6 mg dose that were before all 3 mg comparisons in the testing hierarchy. However, study 503 in elderly patients provides some evidence of efficacy of the 3 mg dose compared to placebo for sleep maintenance as measured by objective WASO. The 1 mg dose in study 503 has the same problem as the 3 mg dose in study 501, i.e., it is below some insignificant 3 mg comparisons in the testing hierarchy. In addition, the differences between 1 mg and placebo in terms of objective WASO at intermediate times between night 1 and night 85 were not nominally significant. Study 509 in the elderly only included subjective measures. The sponsor actually specified the Total Sleep Time as the primary measure in study 509 but if we believe the WASO to be a better measure we can also examine it because the Total Sleep Time results were statistically significant in favor of the Doxepin 6 mg group. The subjective WASO results were also positive. Therefore, study 509 provides some evidence of the superior efficacy of 6 mg over placebo for sleep maintenance as measured by the subjective Total Sleep Time or subjective WASO. Based on these considerations, overall, it seems that there may be sufficient evidence to support the efficacy of the 3 and 6 mg doses for sleep maintenance.

2 INTRODUCTION

2.1 Overview

The IND number for the development of this drug is 67,162.

Five polysomnography (PSG) studies collected efficacy data using 8-hour PSG recordings (objective data) and a morning sleep questionnaire (subjective data) completed by subjects in the sleep laboratory (SP-0401, SP-0402, SP-0501, SP-0503, and SP-0502). Additionally, SP-0503 collected subjective data at home via an Interactive Voice Response System (IVRS). Study SP-0509 was conducted in an outpatient setting and collected only subjective data via an IVRS. The five studies conducted in either adult or elderly subjects with chronic insomnia were designed primarily to evaluate the effects of doxepin on improving sleep maintenance, whereas the transient insomnia study, SP-0502, was designed primarily to assess sleep onset. The majority of subjects enrolled in these studies were female, consistent with the demographics of the chronic insomnia population.

The 4-period crossover, in-patient, PSG Phase 2 studies, SP-0401 (adults) and SP-0402

(elderly), were identical in study design and examined doxepin 1 mg, 3 mg, and 6 mg doses, as well as placebo.

PSG study, SP-0502, was conducted in healthy adult subjects with transient insomnia. This was the only study designed primarily to detect a difference in sleep onset between doxepin and placebo. This study assessed the efficacy of doxepin 6 mg relative to placebo. A laboratory adaptation model (i.e., first night effect) combined with a 3-hour phase advance (early bedtime, 3 hours before usual bedtime) was implemented to induce transient insomnia.

Relevant Meeting Minutes

Excerpts from the 05/31/06 Pre-NDA Meeting Minutes:

The sponsor was advised to present data from the objective studies that would include hour-by-hour calculations of 1) total wake time (TWT); and 2) number of awakenings after sleep onset (NAASO) at each of the visits where assessed.

There was considerable discussion about the statistical analysis of the primary and secondary endpoint data. In terms of the Statistical Analysis Plan (SAP), the Division advised the sponsor to perform a sequential analysis demonstrating effect for both primary and secondary endpoints, with subjective and objective measures for sleep latency and sleep maintenance, by dose, at Week 1 and at end of treatment. Such a sequential analysis would require ranking of the endpoints, such that the sequential analysis would end when one of the endpoints failed to reach statistical significance. The objective endpoints must be considered prior to the analysis of the subjective endpoints, e.g., the hierarchy would analyze Objective WASO, Objective LPS followed by Subjective WASO, Subjective LPS.

We remind you that our statistical analyses will consider persistence of effect on sleep initiation and/or sleep maintenance as a key secondary outcome in your objective studies.

Action Items:

The sponsor will design the clinical study to include an outcome measure of the subject's overall assessment of whether the drug works (i.e., an effect on sleep initiation and/or maintenance).

The sponsor will revise the primary analysis on the primary and secondary endpoints to specify a rank-ordered analysis that will proceed until one of the endpoints fails to reach statistical significance.

2.2 Data Sources

At the time of review the data from the clinical trials were located in the following directory:
<\\cdsesub1\EVSPROD\NDA022036\0000\m5\datasets>

The sponsor's study reports were located in the following directory:

<\\cdsesub1\EVSPROD\NDA022036\0000\m5\53-clin-stud-rep\535-rep-ffic-safety-stud\treatment-of-insomnia\5351-stud-rep-contr>

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study 401

The first subject was enrolled on 27 July 2004 and the last subject completed on 29 September 2004. The study was conducted in eleven study centers located in the United States.

This was a Phase II, randomized, multi-center, double-blind, placebo-controlled, four period crossover, dose-response study. This study was to assess the effects of three doses of doxepin HCl, 1 mg, 3mg, and 6mg, and placebo (each administered for two consecutive nights) in patients (aged 18-64) with primary sleep maintenance insomnia. Patients were to be screened and asked to complete a sleep diary for seven days before the Screening Polysomnography (PSG) Period. During the Screening PSG Period, patients were to receive two consecutive nights of single-blind placebo, followed by eight hours of nightly PSG recording in a sleep center. Patients were to be allowed to leave the sleep center during the day. A 5- or 12- day drug free interval was to separate the Screening PSG Period from randomization and double blind treatment. Patients were to be randomized to a treatment sequence using a Latin square design. Patients were to receive one of four treatments (doxepin HCl 1 mg, 3 mg, 6 mg or placebo) in each of four Treatment Periods using a crossover study design. Each Treatment Period was to be separated by a 5- or 12-day drug-free interval.

The primary efficacy variable is wake time during sleep (WTDS), as determined by PSG assessment. The secondary efficacy variables include wake after sleep onset (WASO), sleep efficiency (SE), total sleep time (TST), latency to persistent sleep (LPS) number of awakenings after sleep onset (NAASO), wake time after sleep (WTAS), and sleep architecture.

EFFICACY ANALYSES

The primary efficacy analysis was to be based on the per-protocol analysis set. Secondary analyses were also to be performed on the ITT analysis set.

Multiplicity Adjustments

For the primary efficacy analysis of WTDS, each of the 3 pairwise comparisons of doxepin HCl treatments to placebo was to be performed using Dunnett's test. No other adjustments were to be made.

Primary Efficacy Analyses

The primary efficacy variable is WTDS. Measures taken from Nights 1 and 2 from the same Treatment Period were to be averaged for analysis. Differences among treatments were to be analyzed using a repeated-measures analysis of variance (ANOVA) model with terms for sequence, patient within sequence, treatment, and period using SAS PROC MIXED. The covariance among the repeated measures was to be modeled separately as unstructured (UN), compound symmetric (CS), and autoregressive (AR(1)). The covariance structure corresponding to the model with the smallest AIC was to be selected for use. A random statement for patient

within sequence was to be included for the AR(1) structure. Pairwise comparisons of each active treatment versus placebo were to be performed using Dunnett's test.

Each endpoint (e.g., the mean value for the two nightly PSG measurements) was to be calculated based on the available data point(s). If both measurements are missing, the value was to be treated as missing, and this missing value was not to be imputed.

All secondary efficacy parameters were to be analyzed using the analysis of variance (ANOVA) model described for WTDS. LPS, latency to REM sleep, and LSO, which are expected to be log-normally distributed, were to be logged prior to analysis. Logging was to be performed prior to averaging values from Nights 1 and 2. Any zero values were to be set to one epoch to permit logging.

3.1.1.1 Sponsor's Results

Eleven investigational sites were initiated and enrolled patients in the study. Study enrollment by center ranged from 2 to 11 patients. Of the 67 patients enrolled, 16 patients were randomized to the 6,P,1,3 treatment sequence, 18 patients were randomized to the P,6,3,1 treatment sequence, 17 patients were randomized to the 3,1,P,6 treatment sequence, and 16 patients were randomized to the 1,3,6,P treatment sequence (1=1 mg doxepin, 3=3 mg doxepin, 6=6 mg doxepin, and P=placebo).

Efficacy analyses were conducted on both the PP and ITT analysis sets. Of the 67 randomized patients, 67 (100%) were included in the ITT and 61 (91%) were included in the PP analysis set. The number of patients randomized to each treatment sequence was similar in each analysis set.

Overall, the mean age was 42.4 years and ranged from 21 to 63 years. Seventy percent of the patients were female. Forty-five percent of the patients were White, 31% were Black or African-American, and 22% were Hispanic. Baseline sleep characteristics were similar across treatment sequence assignments.

Four patients (6%) were excluded from the PP analysis set because they did not meet the WTDS entry criteria. One patient (1%) was excluded from the PP analysis set for not completing all Treatment Periods (02-022, terminated after Treatment Period 1, Night 1), and one additional patient (1%) was excluded due to protocol non-compliance (06-014, excessive napping during Treatment Period 1, Night 2, missed Treatment Period 2, Night 2 due to prohibited medication, and positive urine drug screen at Treatment Period 4, Night 1).

Table 3 presents the sponsor's results for the prespecified primary analysis of the primary endpoint, Wake time during sleep. The sponsor had specified the per-protocol population as the primary analysis population.

Table 3 Study 401: Analyses of Wake Time During Sleep

Analysis Set/Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=61)				
WTDS (minutes) ^[1]				
N	61	61	61	61
Mean (SD)	51.9 (42.25)	43.2 (28.21)	33.4 (21.87)	36.3 (25.17)
Median	38.8	38.0	27.0	28.5
Min. Max	6.3, 243.8	5.0, 142.8	4.5, 111.5	3.3, 152.3
P-Value ^[2]		0.1273	<0.0001	0.0002
Intent-to-treat (N=67)				
WTDS (minutes) ^[1]				
N	66	66	66	67
Mean (SD)	51.5 (40.97)	42.8 (27.48)	34.0 (21.87)	35.8 (24.27)
Median	39.4	37.9	27.9	28.5
Min. Max	6.3, 243.8	5.0, 142.8	4.5, 111.5	3.3, 152.3
P-Value ^[2]		0.0918	<0.0001	<0.0001

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test.

Although the sponsor specified WTDS as the primary efficacy measure the Neurology division usually recommends the WASO endpoint, which the sponsor did also measure in the study (Table 4). The sponsor found that WASO was significantly reduced at the doxepin 1 mg (p=0.0130), 3 mg (p<0.0001) and 6 mg (p<0.0001) dose levels compared with placebo. The mean WASO values in the PP analysis set were 62.1 minutes, 47.3 minutes, 38.6 minutes, and 38.8 minutes for the placebo, doxepin 1 mg, 3 mg and 6 mg dose levels, respectively.

Table 4 Study 401: Analyses of Wake Time After Sleep Onset

Analysis Set/ Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=61)				
WASO (minutes) ^[1]				
N	61	61	61	61
Mean (SD)	62.1 (47.23)	47.3 (30.85)	38.6 (26.67)	38.8 (26.05)
Median	47.0	41.3	31.5	31.5
Min. Max	6.8, 268.8	5.0, 155.8	6.0, 114.3	3.3, 152.5
P-Value ^[2]		0.0130	<0.0001	<0.0001
Intent-to-treat (N=67)				
WASO (minutes) ^[1]				
N	66	66	66	67
Mean (SD)	61.1 (45.79)	46.7 (30.01)	38.9 (26.29)	38.1 (25.16)
Median	47.3	39.9	31.6	31.5
Min. Max	6.8, 268.8	5.0, 155.8	6.0, 114.3	3.3, 152.5
P-Value ^[2]		0.0090	<0.0001	<0.0001

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test.

LPS data were log-transformed prior to analysis. LPS was not significantly different for any dose level of doxepin (Section 15.2, Tables 11.1 and 11.2); however, LPS was numerically reduced in

all three doxepin dose levels compared with placebo, most notably at the 1 and 6 mg dose levels. The pattern of the LPS results was similar using the ITT analysis set; LPS was not significantly different for any dose level of doxepin.

Table 5 Study 401: Analyses of Latency to Persistent Sleep

Analysis Set/ Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=61)				
LPS (minutes) ^[1]				
N	61	61	61	61
Mean (SD)	34.3 (22.27)	30.1 (22.22)	30.8 (21.22)	27.9 (19.92)
Median	31.8	23.0	29.5	24.0
Min, Max	3.5, 104.8	3.3, 85.0	2.3, 91.8	2.0, 87.3
P-Value ^[2]		0.1836	0.2783	0.0681
Intent-to-treat (N=67)				
LPS (minutes) ^[1]				
N	66	66	66	67
Mean (SD)	33.0 (22.02)	29.6 (21.71)	30.1 (20.72)	27.3 (19.44)
Median	28.9	22.4	27.1	22.8
Min, Max	3.5, 104.8	3.3, 85.0	2.3, 91.8	2.0, 87.3
P-Value ^[2]		0.2783	0.3829	0.1001

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test. Data were log-transformed prior to analysis.

Analysis of Subjective Efficacy Measures

Subjective WASO (sWASO) was significantly reduced at the doxepin 6 mg dose level (Dunnett's $p=0.0109$) compared with placebo in the per-protocol population but it was not significantly reduced in the ITT population (Dunnett's $p=0.1168$).

Subjective Latency to Sleep Onset (sLSO) results were log-transformed prior to analysis.

Subjective LSO was significantly decreased at the doxepin 6 mg dose level ($p=0.0437$) in the per-protocol population and ($p=0.0244$) in the ITT population.

3.1.1.2 Reviewer's Comments

This reviewer verified the sponsor's analyses of objective WASO. Objective WASO on night 2 only, showed the same significance pattern as for the average of nights 1 and 2 (which was designated as primary), except that the 1 mg vs. placebo comparison was no longer significant ($p=0.6547$). Because we will see later in the review that there are more standard randomized double blind placebo controlled parallel group phase 3 studies that support an effect of Doxepin 6 mg on objective WASO this review of study 401 will focus on the LPS endpoint.

The sponsor presented only the Dunnett's adjusted p-values for the LPS (and other endpoints) based on the prespecified Proc Mixed analysis because that adjustment for multiplicity of testing was specified in the protocol (LPS $p=0.1001$ for 6 mg vs. placebo). If instead a hierarchical closed testing procedure starting with the 6 mg vs. placebo comparison had been specified then

the p value for 6 mg vs. placebo on the LPS endpoint would be 0.0397 (3 mg $p=0.1751$). However, the analysis of only the second nights of each period did not reveal a significant difference between 6 mg and placebo ($p=0.4828$). A simpler analysis which also provides a way to check the Proc Mixed results is to a) compute the difference in means between 6 mg and placebo for each sequence and as well as the associated variance b) average the sequence specific mean differences and combine the variances to obtain a single Z statistic for testing the treatment difference over all sequences. This is a test for treatment difference stratified by sequence. Note that it assumes no carryover as well as no period effects. The unadjusted p-value which results from this approach for assessing the 6 mg vs. placebo difference in LPS is 0.055. This is more evidence that the difference between 6 mg and placebo on LPS is not robustly statistically significant.

This reviewer also noted a statistically significant period effect ($p=0.042$) for the LPS endpoint, corresponding to the prespecified analysis which analyzed the average of nights 1 and 2 for each period. This means that LPS varied significantly by period even if we focus on the data from only one treatment. In the absence of dropouts and also assuming a lack of a treatment by period interaction a period effect is not a serious problem because in theory the Latin square design ensures that each treatment / period combination has the same frequency. However, in the presence of dropouts or a period by treatment interaction the estimates of treatment differences may be biased. Note that only 4 of the 24 possible sequences of 4 treatments were utilized in this study, so a treatment by period interaction can not be ruled out. There were slight differences in the assignment frequency of each sequence as well, which means each treatment/period combination does not have exactly the same frequency. This could cause bias, especially if there is any missing data or a treatment by period interaction. If we just analyze the LPS in the first period there are no significant treatment group differences for LPS (e.g., 6 mg vs. placebo: $p=0.069$; with Dunnett's adjustment $p=0.169$). For Night 1 LPS data only the p-values were: 6 mg $p=0.0357$ (Dunnett's adjusted $p=0.0911$), 3 mg $p=0.4211$. Also, site 11 ($N=5$) had a bigger treatment group difference for the 6 mg vs. placebo comparison than any other site and excluding this site leads to a loss of significance of the overall 6 mg vs. placebo comparison ($p=0.1325$).

The subjective LSO also had a statistically significant period effect ($p=0.042$). The p-values for sLSO in the ITT population were 0.0090, 0.0357, and 0.0815 for 6 mg, 3 mg, and 1 mg versus placebo, respectively (unadjusted for multiplicity). The Dunnett adjusted p-values were 0.0244, 0.0905, 0.1944, respectively. If we look at night 2 data only instead of averaging nights 1 and 2 then the comparisons with placebo for LSO are no longer significant (e.g., 6 mg $p=0.2601$). Note that night 1 LSO data showed the same significance pattern as the average of nights 1 and 2 data.

In summary, the primary endpoint was met but there are shortcomings of the study which make the latency to persistent sleep results unconvincing. These include a) significant period effect and b) insignificance after applying the prespecified Dunnett's multiplicity adjustment and c) there was no prespecified hierarchy for the secondary endpoints, which include LPS.

3.1.2 Study 402

Study 402 had the same crossover design and analysis as study 401. The only difference was that the patient population was age 65 and up.

The first patient was randomized on 19 September 2004 and the end of the study was 03 January 2005. The study was conducted in eleven study centers located in the United States.

3.1.2.1 Patient Disposition and Baseline Demographics

Eleven investigational sites were initiated and enrolled patients in the study. Study enrollment by center ranged from one to 19 patients. Of the 76 patients enrolled, 18 patients were randomized to the 6,P,1,3 treatment sequence, 22 patients were randomized to the P,6,3,1 treatment sequence, 15 patients were randomized to the 3,1,P,6 treatment sequence, and 21 patients were randomized to the 1,3,6,P treatment sequence (1=1 mg doxepin, 3=3 mg doxepin, 6=6 mg doxepin, and P=placebo).

Seventy-three patients (96%) completed all required Treatment Periods. One patient discontinued the study due to an adverse event during Treatment Period 1. Two patients discontinued the study after Treatment Periods 1 and 2 respectively, due to consent withdrawal.

Overall, the mean age was 71 years and ranged from 64 to 83 years. Sixty-one percent of the patients were female. Eighty-six percent of the patients were White, 11% were Black or African-American, and 3% were Hispanic. Baseline sleep characteristics were similar across treatment sequence assignments. Overall, the mean lights-out time was 22:30, with a range between 21:00 and 24:00. The mean time for patients to fall asleep was 50.2 minutes, with a range between 20 and 120 minutes. The mean total sleep time was 294.5 minutes with a range between 120 and 390 minutes.

3.1.2.2 Sponsor's Results

WTDS was statistically significantly reduced at the doxepin 1 mg ($p=0.0001$), 3 mg ($p<0.0001$), and 6 mg ($p<0.0001$) dose levels compared with placebo in the PP analysis set. The mean WTDS values were 86.0 minutes, 70.1 minutes, 66.4 minutes, and 60.2 minutes for the placebo, doxepin 1 mg, 3 mg and 6 mg dose levels, respectively. The pattern of the WTDS results was similar using the ITT analysis set; WTDS was statistically significantly reduced at all doxepin dose levels ($p<0.0001$).

Table 6 Study 402: Analyses of Wake Time During Sleep

Analysis Set/Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=71)				
WTDS (minutes) ^[1]				
Mean (SD)	86.0 (38.15)	70.1 (32.78)	66.4 (31.56)	60.2 (28.0)
Median	87.3	63.5	58.5	58.8
Min, Max	26.3, 164.8.8	21.8, 195.0	18.3, 167.0	14.8, 150.8
P-Value ^[2]		0.0001	<0.0001	<0.0001
Intent-to-treat (N=76)				
WTDS (minutes) ^[1]				
N	73	74	75	74
Mean (SD)	85.8 (38.39)	69.6 (32.61)	64.8 (31.96)	59.5 (28.3)
Median	87.3	61.4	57.8	55.3
Min, Max	26.3, 164.8	21.8, 195.0	11.8, 167.0	10.5, 150.8
P-Value ^[2]		<0.0001	<0.0001	<0.0001

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test.

WASO was statistically significantly reduced at the doxepin 1 mg ($p<0.0001$), 3 mg ($p<0.0001$), and 6 mg ($p<0.0001$) dose levels compared with placebo in the ITT and PP analysis sets.

Table 7 Study 402: Analyses of Wake Time After Sleep Onset

Analysis Set/ Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=71)				
WASO (minutes) ^[1]				
Mean (SD)	99.0 (40.22)	80.5 (33.97)	72.3 (33.64)	65.2 (29.52)
Median	101.3	79.5	65.0	59.5
Min, Max	30.0, 206.3	23.5, 200.5	18.3, 167.0	15.0, 158.8
P-Value ^[2]		<0.0001	<0.0001	<0.0001
Intent-to-treat (N=76)				
WASO (minutes) ^[1]				
N	73	74	75	74
Mean (SD)	98.0 (41.67)	80.1 (34.26)	70.8 (34.37)	64.3 (29.84)
Median	97.5	78.3	64.0	59.3
Min, Max	30.0, 206.3	23.5, 200.5	11.8, 167.0	12.0, 158.8
P-Value ^[2]		<0.0001	<0.0001	<0.0001

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test.

LPS was not statistically significantly different for any dose level of doxepin (Section 15.2, Tables 11.1 and 11.2); however, LPS was numerically reduced at the 3 mg and 6 mg dose levels compared with placebo.

Table 8 Study 402: Analyses of Latency to Persistent Sleep

Analysis Set/ Parameter	Placebo	Doxepin 1 mg	Doxepin 3 mg	Doxepin 6 mg
Per-Protocol (N=71)				
LPS (minutes) ^[1]				
Mean (SD)	27.1 (19.47)	28.3 (21.34)	23.7 (17.48)	22.4 (14.28)
Median	21.3	23.3	20.3	21.0
Min, Max	4.8, 106.0	3.8, 115.3	2.3, 80.3	2.8, 74.3
P-Value ^[2]		0.9896	0.0964	0.1959
Intent-to-treat (N=76)				
LPS (minutes) ^[1]				
N	73	74	75	74
Mean (SD)	26.8 (19.29)	28.0 (21.01)	23.2 (17.21)	22.4 (14.04)
Median	21.3	23.1	20.0	21.3
Min, Max	4.8, 106.0	3.8, 115.3	2.3, 80.3	2.8, 74.3
P-Value ^[2]		0.9811	0.0667	0.2486

[1]: Measurements taken from Night 1 and Night 2 were averaged. If one of the nights had a missing value, the non-missing value was used.

[2]: P-value comparing each active treatment versus placebo using Dunnett's test. Data were log-transformed prior to analysis.

Sponsor's Analysis of Subjective Efficacy

Subjective WASO (sWASO) was statistically significantly decreased at all doxepin dose levels (1 mg, p=0.0297; 3 mg, p=0.0144; 6 mg, p=0.0074) compared with placebo.

sTST was statistically significantly increased at all doxepin dose levels (1 mg, p=0.0182; 3 mg, p=0.0005; 6 mg, p<0.0001) compared with placebo.

LSO was statistically significantly decreased at the doxepin 6 mg dose level (p=0.0174), and numerically decreased at the 1 mg and 3 mg dose levels compared with placebo.

Reviewer's Comments

The dropouts in this study [1(6%) in the 6/P/1/3 sequence and 2 (13%) in the 3/1/P/6 sequence] raise questions about the validity of the analysis because of the crossover design. Dropouts in a crossover study are more likely to lead to biased results than in the standard parallel group design, for example, if there is a period effect or a treatment by period effect.

This reviewer verified the sponsor's analyses of objective WASO.

This reviewer also noted a significant period effect p=0.0145 on the subjective WASO.

If we look at all periods but Night 2 data only instead of averaging nights 1 and 2 then the comparisons with placebo for subjective WASO are no longer significant (e.g., 6 mg p=0.3945). On the other hand, the night 1 only subjective WASO data were consistent with the results for the average of nights 1 and 2.

The unadjusted p-value for the 6 mg vs. placebo comparison of objective LPS was not significant, p=0.1063. The unadjusted p-value for the 6 mg versus placebo comparison of the subjective LSO is 0.0066 (Dunnett's adjusted p=0.0181). However, since the objective LPS was not statistically significant and it is higher in the Division's preferred testing hierarchy we can not consider the subjective LSO result as significant after adjusting for multiple endpoints. Note that the sponsor did not prespecify a testing hierarchy for the secondary endpoints.

3.1.3 Study 501

The first subject was enrolled on 01 June 2005 and the last subject completed on 31 December 2005. All investigators in this study were based in the United States.

Objectives:

Primary Objective: To evaluate the sedative-hypnotic efficacy of two dose levels of doxepin HCl (doxepin) relative to placebo.

Secondary Objectives: (1) To evaluate the efficacy and safety of two dose levels of doxepin when administered for 35 consecutive nights, and (2) to evaluate the potential for rebound insomnia and withdrawal effects following discontinuation of doxepin after 35 consecutive nights of treatment.

This was a randomized, double-blind, placebo-controlled, parallel-group study designed to assess the efficacy and safety of two dose levels of doxepin, 3 mg and 6 mg, in subjects with primary insomnia and sleep maintenance difficulties.

Baseline

During Visit 3 (Nights -6 and -5), subjects participated in 2 consecutive nights of 8-hour continuous PSG recordings in the sleep center. Subjects who remained eligible for study entry were assigned to one of three treatment groups (placebo, doxepin 3 mg, or doxepin 6 mg) according to a computer-generated randomization scheme. Subjects continued to take single-blind placebo for 5 consecutive nights at home (Nights -4 through 0), until the start of double-blind treatment (Night 1).

Double-blind Treatment Period

During the Double-blind Treatment Period, subjects began 35 consecutive nights of treatment that included supervised administration of study drug during Visits 4, 5, and 6 at the study center and self-administration of study drug at home between visits. During each scheduled study visit, subjects participated in 2 consecutive nights of continuous 8-hour PSG recordings in the sleep center (Nights 1 and 2; Nights 15 and 16; and Nights 29 and 30). After completing each study visit, subjects were dispensed double-blind study drug to self-administer at home (Nights 3 through 14; Nights 17 through 28; and Nights 31 through 35).

The primary efficacy variable was WASO on Night 1.

Sample Size Determination

A total of 240 evaluable patients were planned for this study. Using the phase II study results, a mean difference of 17 minutes between doxepin HCl and placebo and a pooled standard deviation of 33 minutes provide an estimated standardized effect of 0.515. Results from a simulation program using Dunnett's test with an overall alpha level of 5% demonstrate that 80 patients per arm (2 active treatment arms and a control arm) would provide greater than 90% power to detect a significant difference between at least one of the drug treatment arms and the control arm. In order to compensate for patients who fail to qualify for the evaluable population, 282 patients were to be randomized.

Analysis Population

Patients who are randomized to double-blind study medication but who never receive double-blind study drug will not be included in any analysis set, but will be included in selected tabulations based on all randomized patients. All efficacy analyses will be performed on the Intent-to-treat (ITT) analysis set. The ITT analysis set will include all randomized patients who have a corresponding PSG efficacy assessment of WASO at Visit 4, Night 1. Data will be analyzed as randomized and based on observed cases.

Primary Analysis

The primary efficacy variable is WASO at Visit 4, Night 1. Hypothesis tests for the comparison of doxepin HCl at 3 mg and 6 mg to placebo will be analyzed using an analysis of covariance (ANCOVA) model that includes the main effects for treatment and center with the baseline WASO as a covariate. Low enrolling centers may be collapsed for analysis purposes, with sites with fewer than 5 patients being pooled together to form a pseudo-site. Baseline WASO is defined as the mean of Nights -6 and -5. Pairwise comparisons of each active doxepin HCl treatment group versus placebo will be performed using Dunnett's test. The residuals of the model will be examined to determine whether substantial departures from normality are apparent, using the Shapiro-Wilk test. If the data are inconsistent with the assumptions of ANCOVA, sensitivity analyses using an appropriate transformation of the data may be performed or appropriate non-parametric tests may be used to evaluate the differences among treatment groups.

Secondary Analyses

Key secondary variables are WASO, WTDS, TST, SE (whole night), and LPS.

Continuous secondary efficacy variables will be analyzed using the same methods used to compare the WASO values using an ANCOVA model. In addition, data obtained on Night 1 of Visits 4, 5, and 6 will be averaged and then analyzed using the same methods above for the following efficacy variables: WASO, WTDS, TST, SE, LPS, NAASO, sWASO, sTST, LSO, sNAASO, and sleep quality (obtained from the morning questionnaire).

For endpoints that are measured on the two consecutive PSG nights of Visits 4, 5, and 6, the analysis of each endpoint will be based on the Night 1 value. However, supplemental analyses may be performed using the mean of both nights. Supplemental analyses of the key secondary endpoints (WASO, WTDS, TST, SE, and LPS) will be performed using the mean of both nights unless noted otherwise.

LPS, latency to REM sleep, latency to stage 2 sleep, and LSO, which are expected to be log-normally distributed, will be transformed prior to analysis by taking the natural logarithm. The log-transformation will be performed after averaging values from Nights 1 and 2 if average values are used. Any values of zero will be set to 0.5 epoch to permit calculation of the log-transformation.

Multiplicity Adjustments

For the primary efficacy analysis of WASO and analyses of continuous secondary efficacy endpoints, each of the pairwise comparisons of Doxepin HCl treatments to placebo will be

performed using Dunnett's test. For analyses of categorical secondary efficacy endpoints (e.g., CGI assessed for therapeutic effect), each treatment group will be compared to the placebo group using a Bonferroni adjustment. No other adjustments will be made.

Handling Missing Data

For the primary and secondary efficacy analyses, all data will be analyzed using observed values only; missing data will not be imputed. For the total ISI score, if any of the items contributing to the ISI is missing, the total score will be missing as well. The extent of missing data will be assessed during a blinded review of the data, and if appropriate, sensitivity analyses that impute missing data will be performed. Any methods for imputing missing data will be documented prior to unblinding.

Pre-NDA meeting

On 06 March 2006 through 09 March 2006, Somaxon reviewed the final blinded data to identify any remaining data issues that needed to be addressed prior to unblinding the study. The final statistical analysis plan, dated 24 March 2006, was approved prior to unblinding of any study data. On 07 April 2006, Somaxon reviewed the unblinded tables, figures, and listings (TFLs) and identified some formatting changes to the TFLs. Somaxon received the full set of final, unblinded TFLs on 20 May 2006. On 31 May 2006, Somaxon met with the FDA division of Neurology Products to discuss the proposed content of the 505 (b) (2) NDA submission of doxepin I-ICI 1, 3, and 6 mg tablets for the treatment of insomnia. As a result of this pre-NDA meeting, the FDA requested several changes to the planned efficacy presentations for Somaxon Protocol SP-0501 (Study SP-0501). Specifically, the FDA requested that results from the primary and key secondary endpoints be interpreted using a closed system step-down procedure; furthermore, the FDA commented that pairwise comparisons of each treatment group to placebo using Dunnett's test was not necessary.

At the pre-NDA meeting, the FDA reviewers requested that Somaxon implement a procedure that would control for multiple comparisons not only across parameters, but also across timepoints. Specifically, the reviewers requested that Somaxon present results at Visit 4, Night 1 (the pre-specified primary time for analysis), and also at Visit 6, Night 29 to show duration of effect over the course of the study. The reviewers noted that Somaxon would need to implement a method for controlling the overall Type 1 error rate for the multiple analyses; they further noted that no additional adjustment for multiple comparisons across dose groups was needed. Somaxon commented at the meeting that the data had already been analyzed according to the final SAP, and asked for suggestions regarding how to implement such a procedure since Somaxon had already reviewed the unblinded results.

Additional Analyses Recommended by FDA at pre-NDA meeting

The FDA suggested that Somaxon follow a closed-system step-down procedure for interpreting the study results. The step-down system specifies a single comparison at each level (specified below), starting with the comparison of the doxepin 6 mg and placebo groups with respect to the primary endpoint (WASO at Night 1). If the resulting p-value is <0.05 , interpretation of the statistical significance of the next comparison can be made. The procedure stops once a non-significant p-value is reached. Somaxon and the FDA agreed that the objective measures of sleep maintenance and onset would be tested first, followed by subjective measures of sleep

maintenance and onset. Furthermore, the procedure would entail comparing the doxepin 6 mg group to the placebo group; comparisons of the doxepin 3 mg group to placebo would follow in the same order if all comparisons of the doxepin 6 mg group and placebo show statistical significance. It is important to note that while the accepted measure of objective sleep maintenance is WASO, subjective total sleep time (sTST) is the preferred measure of subjective sleep maintenance. Latency to persistence sleep (LPS) is the preferred measure of objective sleep onset, and latency to sleep onset (LSO) is the preferred measure of subjective sleep onset. The following list summarizes the order of the comparisons to be made under this amended analysis plan:

- WASO at Night 1
- WASO at Night 29
- LPS at Night 1
- LPS at Night 29
- sTST at Night 1
- sTST at Night 29

- LSO at Night 1
- LSO at Night 29

Based on these FDA recommendations, all efficacy analyses that were originally conducted using the ITT population will be re-analyzed using an analysis of covariance (ANCOVA) model that includes the main effects for treatment and center with the baseline value as a covariate. Results will be reported for the ITT population only. Sites with fewer than five patients in the ITT population will be pooled into a pseudo-site for analysis. Pairwise comparisons of each active doxepin treatment group versus placebo will be performed within the context of the ANCOVA model; no adjustment for comparing multiple dose groups to the placebo group will be made.

Eight randomized subjects were not included in the Safety Analysis Set because they did not receive at least one dose of double-blind study drug. These eight subjects plus one additional subject (Subject 05-3068 who was randomized to the placebo group but did not have an evaluable PSG recording on Night 1) were not included in the ITT Analysis Set.

The FDA requested that total wake time (TWT) and number of awakenings after sleep onset (NAASO) be presented by hour. The following rules will be applied when calculating TWT and NAASO by hour:

- TWT in the nth hour will be calculated as the (number of wake epochs during the 120 epochs of the nth hour)/2.
- NAASO in the nth hour will be calculated by first determining the NAASO over the entire recording. Each awakening will then be assigned to the hour in which the awakening began. NAASO in the nth hour will be censored (i.e., missing) for hours that precede sleep onset, as defined by 20 consecutive epochs of sleep. NAASO in the nth hour will be defined to be zero for hours in which no awakening began. These parameters will be analyzed using the ITT population, using an ANCOVA model that includes the main effects for treatment and center with the baseline value as a covariate. Sites with fewer than five patients in the ITT population will be pooled into a pseudo-site for analysis. Pairwise comparisons of each active doxepin

treatment group versus placebo will be performed within the context of the ANCOVA model; no adjustment for multiple comparisons will be made.

3.1.3.1 Patient Disposition

A total of 1,082 subjects were screened for this study. A summary of disposition for all randomized subjects is provided in Table 9. Of the 229 randomized subjects, 221 subjects (97%) received double-blind study drug and were included in the Safety Analysis Set. Overall, 203 subjects (89%) completed the study. Of the 26 subjects (11%) who withdrew from the study, eight subjects (3%) withdrew after randomization but before receiving a single dose of double-blind study drug, and 18 subjects (8%) withdrew during the double-blind Treatment Period. Six randomized subjects withdrew from the study due to an AE. Two of these subjects (one in the placebo group and one in the doxepin 3 mg group) did not receive a dose of double-blind study drug. The remaining four subjects (one in the doxepin 3 mg group and three in the doxepin 6 mg group) withdrew from the study during the Double-blind Treatment Period. Subject 06-3223 withdrew due to hypertension that began during the Placebo Lead-in Period, prior to double-blind treatment. There were no deaths reported during the study or 30 days following administration of the last dose of study drug.

Table 9 Study 501 Patient Disposition

Disposition	Placebo (N=76)	Doxepin 3 mg (N=77)	Doxepin 6 mg (N=76)	Total (N=229)
Completed the Study	67 (88%)	68 (88%)	68 (89%)	203 (89%)
Withdrew from the Study	9 (12%)	9 (12%)	8 (11%)	26 (11%)
Adverse Event	1 (1%)	2 (3%)	3 (4%)	6 (3%)
Protocol Violation	0 (0%)	1 (1%)	0 (0%)	1 (<1%)
Noncompliance	2 (3%)	2 (3%)	2 (3%)	6 (3%)
Consent Withdrawn	3 (4%)	2 (3%)	0 (0%)	5 (2%)
Lost to Follow-up	1 (1%)	0 (0%)	1 (1%)	2 (1%)
Death	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other	2 (3%)	2 (3%)	2 (3%)	6 (3%)
Received Double-blind Study Drug¹	73 (96%)	75 (97%)	73 (96%)	221 (97%)
Completed Double-blind Treatment Period	67 (88%)	68 (88%)	68 (89%)	203 (89%)
Withdrew During Double-blind Treatment Period	6 (8%)	7 (9%)	5 (7%)	18 (8%)
Adverse Event	0 (0%)	1 (1%)	3 (4%)	4 (2%)
Protocol Violation	0 (0%)	1 (1%)	0 (0%)	1 (<1%)
Noncompliance	2 (3%)	2 (3%)	2 (3%)	6 (3%)
Consent Withdrawn	2 (3%)	1 (1%)	0 (0%)	3 (1%)
Lost to Follow-up	1 (1%)	0 (0%)	0 (0%)	1 (<1%)
Death	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other	1 (1%)	2 (3%)	0 (0%)	3 (1%)
Received Single-blind Placebo Discontinuation Study Drug	67 (88%)	68 (88%)	68 (89%)	203 (89%)
Completed Discontinuation Period	67 (88%)	68 (88%)	68 (89%)	203 (89%)

¹ Eight randomized subjects (three in the placebo group, two in the doxepin 3 mg group, and three in the doxepin 6 mg group) withdrew from the study before receiving a dose of double-blind study drug.

Approximately 12% of subjects in each treatment group withdrew from the study. There were no important differences across treatment groups regarding the reasons for study withdrawal. During the Double-blind Treatment Period, the most frequent reasons for study withdrawal were noncompliance (3%) and consent withdrawn (3%) for subjects in the placebo group; noncompliance (3%) and other (3%) for subjects in the doxepin 3 mg group; and noncompliance (3%) and AEs (4%) for subjects in the doxepin 6 mg group. Three of the six subjects who were withdrawn due to noncompliance were participating in this study at more than one study center. One subject in the doxepin 6 mg group, Subject 06-3228, was discontinued after receiving approximately 14 days of double-blind study drug once it was discovered that she previously had completed the study as Subject 19-3112 at another study center.

Twenty-two of the 24 study centers randomized subjects into the study. Data from six centers (Nos. 9, 15, 16, 18, 25, and 26) with low enrollment (fewer than five subjects in the ITT Analysis Set) were pooled to form one pseudo-center, as described in the Statistical Analysis Plan (SAP).

3.1.3.2 Baseline Demographic and Disease Characteristics

Demographic and other baseline characteristics were similar across treatment groups, as summarized in Table 10. The majority of the study population was female: female (73%) and male (27%). The mean age was 44.5 years. Subjects ethnicities were White (48%), Black/African American (33%), Hispanic (16%), Asian (1%), and Other (2%).

Table 10 Study 501: Demographic and Other Baseline Characteristics: Safety Analysis Set

Variable/Category	Placebo (N=73)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)	Total (N=221)
Age (years)				
Mean (SD)	43.6 (12.31)	45.5 (10.56)	44.2 (11.05)	44.5 (11.30)
Range	(18–64)	(20–64)	(19–63)	(18–64)
Sex [n (%)]				
Male	22 (30%)	17 (23%)	21 (29%)	60 (27%)
Female	51 (70%)	58 (77%)	52 (71%)	161 (73%)
Race/Ethnicity [n (%)]				
White	35 (48%)	33 (44%)	39 (53%)	107 (48%)
Black/African American	25 (34%)	26 (35%)	21 (29%)	72 (33%)
Hispanic	11 (15%)	15 (20%)	10 (14%)	36 (16%)
Asian	1 (1%)	1 (1%)	0 (0%)	2 (1%)
Other	1 (1%)	0 (0%)	3 (4%)	4 (2%)
Weight (kg)				
Mean (SD)	74.4 (13.93)	77.5 (14.54)	77.8 (15.35)	76.6 (14.63)
Range	(52–106)	(47–117)	(51–119)	(47–119)
BMI (kg/m²)				
Mean (SD)	26.4 (4.54)	27.8 (4.94)	27.4 (4.14)	27.2 (4.57)
Range	(18.2–40.6)	(19.1–41.5)	(19.2–38.0)	(18.2–41.5)

Copied from page 83 of study report

Eight randomized subjects were not included in the Safety Analysis Set because they did not receive at least one dose of double-blind study drug. These eight subjects plus one additional subject (Subject 05-3068 who was randomized to the placebo group but did not have an evaluable PSG recording on Night 1) were not included in the ITT Analysis Set.

3.1.3.3 Sponsor's Results

Primary Efficacy Variable – Wake After Sleep Onset on Night 1

The primary efficacy variable was WASO on Night 1. Summary statistics for WASO at baseline, Night 1 are presented in Table 11 using the ITT population.

A summary of WASO at baseline and Night 1 by treatment group using the ITT Analysis Set is provided in Table 11. At baseline, the mean WASO, approximately 65 minutes, was similar across the treatment groups. On Night 1, there were statistically significant decreases ($p < 0.0001$) in the mean WASO for each doxepin group, 3 mg and 6 mg, compared with the placebo group. The LS mean WASO at night 1 was shorter for the doxepin 3 mg and 6 mg groups by 26.0 minutes and 30.8 minutes, respectively, compared with the placebo group.

Table 11 Study 501: Primary Efficacy Variable – WASO on Night 1: ITT Analysis Set

WASO (minutes)	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
Baseline (Mean of Nights -6 and -5)			
Mean (SD)	65.6 (37.03)	67.8 (33.56)	65.0 (33.23)
Median (Range)	62.0 (7.0–193.0)	65.3 (9.3–167.5)	58.8 (2.5–178.0)
Night 1 (Visit 4)			
Mean (SD)	66.7 (50.28)	41.4 (31.51)	36.3 (26.14)
Median (Range)	55.5 (6.5–292.5)	32.5 (3.5–175.5)	29.5 (3.0–126.5)
LS Mean (Std. Err.)	69.2 (3.88)	43.2 (3.80)	38.3 (3.81)
Diff. of LS Mean (Std. Err.)		-26.0 (5.18)	-30.8 (5.21)
95% CI of LS Mean Diff.		(-37.5, -14.4)	(-42.4, -19.2)
p-value ¹		$p < 0.0001$	$p < 0.0001$

¹ p-value comparing each active treatment to placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using Dunnett's test.

A summary of objective WASO results for Night 15, Night 29, and the average of Nights 1, 15, and 29 are presented in Table 12.

Table 12 Study 501: WASO on Night 15, Night 29, and the Average of Nights 1, 15, and 29:

WASO (minutes)	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
Night 15 (Visit 5)	n=69	n=69	n=70
Mean (SD)	60.7 (52.87)	44.8 (27.26)	41.9 (29.67)
Median (Range)	45.5 (5.0–300.0)	41.0 (2.0–131.0)	35.0 (3.0–137.5)
Diff. of LS Mean (Std. Err.)		-17.3 (5.69)	-18.6 (5.66)
95% CI of LS Mean Diff.		(-30.0, -4.6)	(-31.2, -6.0)
p-value ¹		p=0.0053	p=0.0023
Night 29 (Visit 6)	n=68	n=68	n=69
Mean (SD)	61.8 (39.71)	47.3 (43.53)	41.2 (37.91)
Median (Range)	53.8 (11.5–171.0)	41.5 (1.5–318.5)	25.0 (5.5–208.0)
Diff. of LS Mean (Std. Err.)		-15.7 (6.44)	-22.3 (6.40)
95% CI of LS Mean Diff.		(-30.0, -1.3)	(-36.6, -8.0)
p-value ¹		p=0.0299	p=0.0012
Average of Nights 1, 15, and 29	n=72	n=75	n=73
Mean (SD)	62.5 (37.67)	44.4 (26.58)	39.5 (26.09)
Median (Range)	56.1 (8.0–179.8)	37.5 (4.5–144.0)	33.2 (4.3–131.5)
Diff. of LS Mean (Std. Err.)		-18.9 (4.07)	-23.4 (4.09)
95% CI of LS Mean Diff.		(-27.9, -9.8)	(-32.5, -14.3)
p-value ¹		p<0.0001	p<0.0001

¹ p-value comparing each active treatment versus placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using Dunnett's test.

Latency to Persistent Sleep

A summary of LPS at baseline, Night 1, Night 15, and Night 29 by treatment group using the ITT Analysis Set is provided in Table 13. At baseline, the mean LPS was similar across the treatment groups. On Night 1, there were statistically significant decreases in the geometric LS mean LPS value for each doxepin group, 3 mg and 6 mg, compared with the placebo group. The geometric LS mean for LPS was 18.1 and 16.7 minutes for the doxepin 3 mg and 6 mg groups, respectively, compared with 26.8 minutes for the placebo group.

At Night 29, improvement in LPS in the doxepin 3 mg and 6 mg groups was comparable to the improvement observed on Night 1 (geometric LS means for LPS on Night 29 were 17.8 and 16.6 minutes, respectively); however, these results were not statistically significant due, in part, to a substantial placebo response.

Results for LPS on the average of Nights 1, 15, and 29 for each doxepin group, 3 mg and 6 mg, compared with placebo were not statistically significant based on the ITT Analysis Set.

*Reviewer's Comment: The log transformation of LPS was specified as the endpoint for the analysis because the log LPS has a distribution that is closer to a normal distribution than the untransformed LPS, which tends to have an asymmetric or skewed distribution. A mean of the log transformed LPS values can be shown by using properties of the logarithm to be equal to the log of the geometric mean, which has the form $(y_1 * y_2 * \dots * y_n)^{1/n}$. Therefore, taking the exponent of the mean of the log transformed values gives the geometric mean which is back on the original scale and thus is more interpretable than the mean of the log transformed values.*

Table 13 Study 501: LPS at Baseline, Night 1, Night 15, Night 29, and Average of Nights 1, 15, and 29: ITT Analysis Set

LPS (minutes)	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
Baseline (Mean of Nights -6 and -5)	n=72	n=75	n=73
Mean (SD)	38.0 (28.56)	35.9 (29.84)	39.1 (34.10)
Geometric Mean	27.6	27.4	25.9
Median (Range)	32.0 (1.8–146.5)	26.8 (5.0–191.5)	30.3 (1.5–194.8)
Night 1 (Visit 4)	n=72	n=75	n=73
Mean (SD)	45.0 (54.91)	26.7 (23.42)	27.1 (25.42)
Median (Range)	28.5 (2.5–394.0)	17.5 (2.5–103.5)	19.5 (0.5–163.5)
Geometric LS Mean ¹	26.8	18.1	16.7
LS Mean Ratio		0.7	0.6
95% CI of LS Mean Ratio		(0.5, 0.9)	(0.5, 0.9)
p-value ²		p=0.0110	p=0.0018
Night 15 (Visit 5)	n=69	n=69	n=70
Mean (SD)	33.2 (39.75)	38.2 (40.52)	31.8 (36.57)
Median (Range)	22.0 (2.0–237.0)	22.5 (1.0–256.0)	17.3 (0.5–170.0)
Geometric LS Mean ¹	19.2	23.5	17.6
LS Mean Ratio		1.2	0.9
95% CI of LS Mean Ratio		(0.8, 1.8)	(0.6, 1.3)
p-value ²		p=0.3644	p=0.8315
Night 29 (Visit 6)	n=68	n=68	n=69
Mean (SD)	31.3 (35.98)	28.0 (25.99)	24.7 (21.48)
Median (Range)	16.5 (0.5–204.0)	20.3 (0.5–130.5)	18.5 (1.0–81.0)
Geometric LS Mean ¹	16.7	17.8	16.6
LS Mean Ratio		1.1	1.0
95% CI of LS Mean Ratio		(0.7, 1.5)	(0.7, 1.4)
p-value ²		p=0.9008	p=0.9989
Average of Nights 1, 15, and 29	n=72	n=75	n=73
Mean (SD)	36.9 (34.05)	31.0 (24.92)	27.8 (21.62)
Median (Range)	27.6 (4.5–165.0)	21.8 (3.5–131.0)	21.3 (1.0–92.2)
Geometric LS Mean ¹	25.1	22.9	20.4
LS Mean Ratio		0.9	0.8
95% CI of LS Mean Ratio		(0.7, 1.2)	(0.6, 1.0)
p-value ²		p=0.6065	p=0.1076

¹ Analysis was performed on log-transformed data. The LS mean values were converted to original scale by taking the anti-log.

² p-value comparing each active treatment to placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using Dunnett's test.

Subjective Efficacy Measures

A summary of the subjective variables sTST, sWASO, and sNAASO at baseline, Day 2, Day 16, and Day 30 is provided in Table 14. Note that for each measure the Day k score (k=2, 16, or 30) was obtained on day k but it pertains to the previous night, Night k-1. In general, at baseline the subjective sleep variables sTST, sWASO, and sNAASO were similar across the treatment groups. Results from the mornings of Day 2, Day 16, and Day 30 are summarized below.

- sTST and sWASO – After Night 1, there were statistically significant improvements in the mean sTST and mean sWASO for each doxepin group, 3 mg and 6 mg, compared with the placebo group. These improvements continued throughout dosing on Night 29 (Day 30 assessment) for each doxepin group, 3 mg and 6 mg. However, the results were not statistically significant due, in part, to a substantial placebo response.

- sNAASO – There was no statistically significant difference in sNAASO for each doxepin group, 3 mg and 6 mg, compared with the placebo group on Day 2, Day 16, or Day 30 (after dosing on Night 1, Night 15, and Night 29).

Table 14 Study 501: Subjective Sleep Variables sTST, sWASO, and sNAASO at Baseline, Day 2, Day 16, and Day 30: ITT Analysis Set

Variable Using the Morning Questionnaire	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
sTST (minutes)	n=72	n=75	n=73
Baseline (Mean of Days -5 and -4) (SD)	341.2 (52.20)	330.0 (63.23)	339.3 (61.45)
Day 2 (Visit 4)	n=72	n=75	n=73
Mean (SD)	348.3 (70.03)	361.8 (64.03)	369.0 (78.39)
p-value [†]		p=0.0169	p=0.0256
Day 16 (Visit 5)	n=70	n=69	n=70
Mean (SD)	353.6 (85.25)	361.0 (67.22)	371.5 (72.06)
p-value [†]		p=0.3260	p=0.1611
Day 30 (Visit 6)	n=68	n=69	n=69
Mean (SD)	365.2 (68.23)	360.7 (68.72)	373.0 (75.30)
p-value [†]		p=0.9972	p=0.6831
sWASO (minutes)	n=72	n=75	n=73
Baseline (Mean of Days -5 and -4) (SD)	74.6 (39.74)	80.6 (48.12)	78.2 (43.06)
Day 2 (Visit 4)	n=71	n=75	n=73
Mean (SD)	72.5 (45.91)	55.7 (39.81)	54.9 (44.76)
p-value [†]		p=0.0005	p=0.0007
Day 16 (Visit 5)	n=70	n=69	n=70
Mean (SD)	66.9 (59.65)	59.0 (54.14)	58.4 (49.09)
p-value [†]		p=0.2582	p=0.3357
Day 30 (Visit 6)	n=68	n=69	n=69
Mean (SD)	59.6 (43.21)	63.1 (47.24)	58.2 (53.07)
p-value [†]		p=0.8958	p=0.8020

A summary of the subjective variables LSO and sleep quality at baseline, Day 2, Day 16, and Day 30 is provided in Table 15. In general, at baseline LSO and sleep quality were similar across the treatment groups. Results from the mornings of Day 2, Day 16, and Day 30 are summarized below.

- LSO – Although there were numerically greater improvements from baseline in the mean LSO after dosing on Night 1 for each doxepin group, the results were not statistically significant. Numerical improvements in sleep onset continued throughout treatment (Day 30 assessment) for each doxepin group, 3 mg and 6 mg, but these results were not statistically significant.

Table 15 Study 501: Subjective Sleep Variables LSO and Sleep Quality at Baseline, Day 2, Day 16, and Day 30: ITT Analysis Set

Variable Using the Morning Questionnaire	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
LSO (minutes)¹	n=72	n=75	n=73
Baseline (Mean of Days -5 and -4) (SD)	54.7 (30.52)	61.4 (39.43)	64.0 (43.45)
Day 2 (Visit 4)	n=72	n=75	n=73
Mean (SD)	56.4 (46.66)	50.3 (36.73)	55.7 (56.96)
p-value ²		p=0.2296	p=0.0960
Day 16 (Visit 5)	n=70	n=69	n=70
Mean (SD)	55.1 (62.42)	52.9 (35.17)	48.7 (42.35)
p-value ²		p=0.9838	p=0.2689
Day 30 (Visit 6)	n=68	n=69	n=68
Mean (SD)	44.1 (41.26)	48.9 (34.71)	48.3 (47.40)
p-value ²		p=0.3567	p=0.8242

All Night Sleep Efficiency

A summary of Sleep Efficiency (SE) overall at baseline, Night 1, Night 15, and Night 29 by treatment group using the ITT Analysis Set is presented in Table 16. At baseline, the mean SE overall was similar across the treatment groups. On Night 1, there were statistically significant improvements in mean SE overall for each doxepin group compared with placebo. The LS mean SE overall was greater for the doxepin 3 mg and 6 mg groups by 8.6% and 9.8%, respectively, compared with the placebo group. There were statistically significant increases in mean SE overall on Night 1 for the doxepin groups compared with the placebo group, which were sustained on Night 15 (6 mg group) and Night 29 (3 mg and 6 mg groups). Additionally, there were statistically significant increases in mean SE overall for the average of Nights 1, 15, and 29 for each doxepin group compared with the placebo group.

Table 16 Study 501: SE All Night Over Time: ITT Analysis Set

Sleep Efficiency Overall (%)	Placebo (N=72)	Doxepin 3 mg (N=75)	Doxepin 6 mg (N=73)
Baseline (Mean of Nights -6 and -5)	n=72	n=75	n=73
Mean (SD)	79.2 (9.31)	79.2 (9.60)	79.2 (8.98)
Median (Range)	80.5 (52.3–95.7)	81.4 (53.0–94.7)	79.4 (56.9–94.1)
Night 1 (Visit 4)	n=72	n=75	n=73
Mean (SD)	77.9 (15.05)	86.5 (8.68)	87.6 (7.72)
Median (Range)	81.1 (14.3–97.6)	89.2 (54.4–97.6)	90.5 (62.9–98.4)
Diff. of LS Mean (Std. Err.)		8.6 (1.46)	9.8 (1.46)
95% CI of LS Mean Diff.		(5.3, 11.8)	(6.6, 13.1)
p-value ¹		p<0.0001	p<0.0001
Night 15 (Visit 5)	n=69	n=69	n=70
Mean (SD)	81.2 (13.42)	83.7 (10.12)	85.7 (10.70)
Median (Range)	83.2 (35.7–97.4)	85.2 (46.4–98.2)	88.1 (46.6–97.8)
Diff. of LS Mean (Std. Err.)		2.6 (1.66)	4.4 (1.65)
95% CI of LS Mean Diff.		(-1.1, 6.4)	(0.7, 8.1)
p-value ¹		p=0.1977	p=0.0157
Night 29 (Visit 6)	n=68	n=68	n=69
Mean (SD)	81.5 (10.52)	85.0 (10.92)	87.3 (9.37)
Median (Range)	82.6 (54.6–96.1)	88.0 (27.5–97.0)	89.8 (52.4–98.4)
Diff. of LS Mean (Std. Err.)		3.8 (1.52)	5.8 (1.51)
95% CI of LS Mean Diff.		(0.4, 7.1)	(2.5, 9.2)
p-value ¹		p=0.0262	p=0.0003
Average of Nights 1, 15, and 29	n=72	n=75	n=73
Mean (SD)	80.2 (11.03)	85.1 (8.95)	86.9 (7.66)
Median (Range)	81.1 (45.0–95.8)	86.6 (50.2–97.6)	88.7 (69.1–96.7)
Diff. of LS Mean (Std. Err.)		4.8 (1.18)	6.8 (1.19)
95% CI of LS Mean Diff.		(2.2, 7.5)	(4.1, 9.4)
p-value ¹		p=0.0001	p<0.0001

¹ p-value comparing each active treatment versus placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using Dunnett's test.

Source: study report page 94

Reviewer's comment: The estimated correlation between the SE overall and the TST is 1.00 because the SE is defined as the TST divided by the time in bed and time in bed was almost always equal to 480 minutes (with a few slight exceptions). Thus, SE all night is redundant if considered after TST.

Sleep Efficiency by hour of the night, compared with placebo, was statistically significantly improved on Night 1 at Hours 2, 3, 6, and 8 for the doxepin 3 mg group, and at Hours 2–8 in the doxepin 6 mg group.

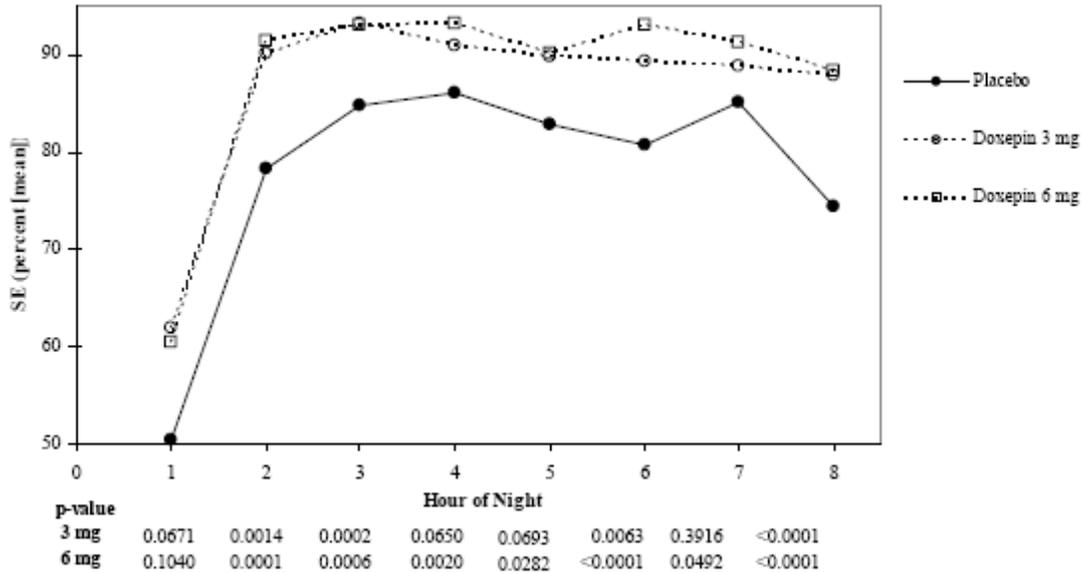


Figure 1 Study 501: SE by Hour of the Night on Night 1: ITT Analysis Set

3.1.3.4 Reviewer’s Results

Objective WASO

This reviewer verified the sponsor’s primary analysis of the WASO at Night 1 which revealed a statistically significant difference between Doxepin 6 mg and placebo, favoring Doxepin. Eight randomized patients had no post-baseline efficacy data and one of these had no baseline data either. Using baseline observation carried forward imputation (BOCF) for these ITT patients with no post baseline efficacy data doesn’t change the significance of the result for WASO at Night 1.

There was also a significant difference between 6 mg Doxepin and placebo on the WASO at Night 29. This was based on an observed cases analysis, i.e., with no imputation for missing data. The result was still significant if missing data were imputed using either a) BOCF (baseline carried forward) or b) LOCF (last observation carried forward) [see Table 17].

Table 17 Comparison of OC and LOCF analyses for objective WASO

POPULATION	NIGHT	PLACEBO		3 MG				6 MG			
		N	MEAN (S.D.)	N	MEAN (S.D.)	Difference from Placebo LSMEAN (s.E.)	p-value	N	MEAN (S.D.)	Difference from Placebo LSMEAN (s.E.)	p-value
OC	Night 15	70	60.8(52.5)	69	44.8(27.3)	-17.3(5.7)	0.0025	70	41.9(29.7)	-18.6(5.6)	0.0011
	Night 16	70	60.7(48.8)	68	43.4(27.1)	-19.6(5.7)	0.0007	68	42.2(31.2)	-19.4(5.7)	0.0008
	Night 29	69	61.6(39.4)	68	47.3(43.5)	-15.4(6.4)	0.0173	69	41.2(37.9)	-22.0(6.4)	0.0007
	Night 30	67	60.3(53.6)	67	44.5(30.2)	-18.6(6.1)	0.0025	68	40.1(31.1)	-22.3(6.0)	0.0003
LOCF	Night 15	73	60.5(51.9)	75	44.4(28.1)	-17.4(5.4)	0.0015	73	41.5(29.1)	-19.1(5.4)	0.0005
	Night 16	73	60.1(48.3)	75	43.4(27.8)	-18.8(5.4)	0.0006	73	42.4(30.4)	-17.7(5.4)	0.0012
	Night 29	73	60.8(38.9)	75	46.7(42.7)	-14.8(6.0)	0.0145	73	40.5(37.1)	-21.4(6.0)	0.0005
	Night 30	73	60.0(51.8)	75	44.2(30.5)	-17.1(5.6)	0.0025	73	39.0(30.5)	-21.1(5.6)	0.0002

Another approach to assessing the impact of missing data on the results is a mixed model for repeated measures. This approach models all of the observed post-baseline WASO scores (first night of each visit) simultaneously. This reviewer’s model included baseline score as a covariate, center effects, treatment group effects, visit (as a class variable which avoids assuming a particular functional relationship for how the WASO changes over time) and effects for the interaction between visit and treatment group. The within subject covariance structure for repeated measures was specified as unstructured (to avoid questionable preconceived notions about how the correlation between two observations on the same subject varies according to the amount of time between them). The analysis of this model agreed with the other three methods considered here. Therefore, it seems relatively unlikely that the dropouts (4 placebo, 7 Doxepin 3 mg, and 4 Doxepin 6 mg) would alter this result if their data was complete.

Subjective WASO

The analyses of the subjective WASO by night for the observed cases are summarized in Table 18. Differences between 6 mg and placebo and 3 mg and placebo on Subjective WASO were only significant at Night 1 (not for Night 29 or night 15). The sponsor specified the first of the two nights for each visit as the timepoint to be tested, rather than the second night or the average. The difference between Doxepin 6 mg and placebo on subjective WASO on night 30 (2nd night of last Visit) was nominally significant (p=0.0009). This difference was also nominally significant at night 16 (p=0.0418 without the Dunnett’s multiplicity adjustment), but not at night 15 (p=0.2016). In summary, if we use the nights specified for testing by the sponsor and we start with the last one and work backwards we have to stop testing the subjective WASO at Night 29 (unadjusted p=0.6282). Therefore, using the hierarchical approach starting at the end with the sponsor’s prespecified Nights for testing we are unable to conclude that there were any significant differences between 6 mg and placebo on the subjective WASO.

Table 18 Analyses of subjective WASO by Night (Observed Cases Analyses)

	N	MEAN (S.D.)	N	MEAN (S.D.)	LS MEAN (S.E.) DIFFERENCE FROM PLACEBO	P- VALUE	N	MEAN (S.D.)	LS MEAN (S.E.) DIFFERENCE FROM PLACEBO	P- VALUE
Baseline	75	73.88(39.66)	76	81.51 (48.46)	8.71	0.209	76	75.66 (44.11)	2.28	0.741
Night 1	72	71.9 (45.9)	75	55.7 (39.8)	-20.6 (5.6)	<0.001	73	54.9 (44.8)	-20.3 (5.6)	<0.001
Night 2	73	63.6 (45.0)	74	60.9 (49.3)	-6.9 (6.0)	0.257	73	49.4 (34.3)	-16.0 (6.0)	0.009
Avg N1,N2	73	67.7 (40.8)	75	58.0 (38.9)	-13.7 (4.8)	0.005	73	52.2 (35.5)	-18.0 (4.8)	<0.001
Night 15	71	66.4 (59.4)	69	59.0 (54.1)	-11.6 (8.0)	0.151	70	58.4 (49.1)	-10.2 (8.0)	0.202
Night 16	70	60.3 (44.7)	69	54.8 (48.1)	-7.5 (6.5)	0.248	68	50.5 (40.8)	-13.3 (6.5)	0.042
Avg N15,16	71	63.2 (43.1)	69	56.9 (44.1)	-9.6 (5.8)	0.103	70	54.0 (39.9)	-11.8 (5.8)	0.042
Night 29	69	59.1 (43.0)	69	63.1 (47.2)	3.2 (7.1)	0.648	69	58.2 (53.1)	-3.4 (7.0)	0.628
Night 30	69	70.7 (56.5)	67	60.6 (45.6)	-13.7 (7.2)	0.059	67	49.7 (43.7)	-24.1 (7.2)	0.001
Avg N29,30	69	64.9 (41.9)	69	61.7 (42.9)	-5.8 (6.1)	0.343	69	53.7 (43.9)	-14.2 (6.1)	0.020

Effect of Missing Subjective WASO data

As can be seen in Table 19 significance conclusions for 6 mg vs. placebo comparisons on subjective WASO based on analyses that used LOCF imputation for those who dropped out but had some post-baseline data were the same as those for observed cases analyses except at Night 16 (OC p=0.042; LOCF p=0.053). Therefore, for the most part, it seems that the dropouts would not have too much effect on the results of the analyses if their data was complete.

Table 19 Study 501: Subjective WASO: Comparison of OC and ITT-LOCF analyses

Population	Night	PLACEBO		3 MG		Difference from Placebo LSMEAN (S.E.)	p-value	6 MG		Difference from Placebo LSMEAN (S.E.)	p-value
		N	MEAN (S.D.)	N	MEAN (S.D.)			N	MEAN (S.D.)		
OC	Night 1	72	71.9 (45.9)	75	55.7 (39.8)	-20.6 (5.6)	0.000	73	54.9 (44.8)	-20.3 (5.6)	0.000
	Night 2	73	63.6 (45.0)	74	60.9 (49.3)	-6.9 (6.0)	0.257	73	49.4 (34.3)	-16.0 (6.0)	0.009
	Night 15	71	66.4 (59.4)	69	59.0 (54.1)	-11.6 (8.0)	0.151	70	58.4 (49.1)	-10.2 (8.0)	0.202
	Night 16	70	60.3 (44.7)	69	54.8 (48.1)	-7.5 (6.5)	0.248	68	50.5 (40.8)	-13.3 (6.5)	0.042
	Night 29	69	59.1 (43.0)	69	63.1 (47.2)	3.2 (7.1)	0.648	69	58.2 (53.1)	-3.4 (7.0)	0.628
	Night 30	69	70.7 (56.5)	67	60.6 (45.6)	-13.7 (7.2)	0.059	67	49.7 (43.7)	-24.1 (7.2)	0.001
	Avg N29,N30	69	64.9 (41.9)	69	61.7 (42.9)	-5.8 (6.1)	0.343	69	53.7 (43.9)	-14.2 (6.1)	0.020
LOCF	Night 15	73	65.2 (59.1)	75	57.8 (52.4)	-11.6 (7.6)	0.1304	74	57.4 (48.0)	-9.9 (7.6)	0.1966
	Night 16	73	59.2 (44.4)	75	52.9 (47.1)	-9.1 (6.3)	0.1472	73	49.2 (39.9)	-12.2 (6.3)	0.0532
	Night 29	73	57.2 (42.8)	75	61.6 (46.0)	3.0 (6.7)	0.6527	74	58.1 (52.0)	-1.1 (6.7)	0.8690
	Night 30	73	69.0 (55.6)	75	57.8 (44.5)	-14.9 (6.8)	0.0291	73	48.2 (42.4)	-23.2 (6.8)	0.0008
	Avg N29, N30	73	63.5 (41.4)	75	59.5 (42.3)	-6.4 (5.8)	0.2685	73	52.4 (43.1)	-13.4 (5.8)	0.0221

Latency to Persistent Sleep

Differences in LPS were not significant compared to placebo for either 3 mg or 6 mg on Night 2 (Dunnett’s adjusted: p=0.523 and p=0.366, respectively). Without Dunnett’s adjustment the p-values were 0.334 and 0.222, respectively. However, both 3 mg and 6 mg were significant compared to placebo for Night 1 alone, as well as for the average of Nights 1 and 2. There were no statistically significant treatment group differences in LPS at later times (Night 15 or Night 29). The analyses of the objective LPS by night for the observed cases are summarized in Table 20.

Table 20 Study 501: Analyses of objective LPS by Night (Observed Cases Analyses)

	N	MEAN OF LOG LPS (S.D.)	GEO-METRIC MEAN OF LPS	N	MEAN OF LOG LPS (S.D.)	GEO-METRIC MEAN OF LPS	P-VALUE FOR DIFF FROM PLACEBO	N	MEAN OF LOG LPS (S.D.)	GEO-METRIC MEAN OF LPS	P-VALUE FOR DIFF FROM PLACEBO
Baseline	76	3.35 (0.89)	28.5	76	3.31 (0.74)	27.3	0.976	76	3.26 (1.00)	26.0	0.542
Night 1	72	3.33 (1.00)	28.0	75	2.93 (0.87)	18.7	0.006	73	2.80 (1.17)	16.5	0.001
Night 2	72	3.05 (1.06)	21.2	74	2.84 (1.01)	17.2	0.282	73	2.77 (1.24)	16.0	0.183
Avg N1,N2	73	3.31 (0.88)	27.3	75	2.98 (0.77)	19.7	0.010	73	2.93 (1.03)	18.7	0.007
Night 15	70	2.98 (1.05)	19.7	69	3.16 (1.07)	23.5	0.227	70	2.84 (1.23)	17.1	0.592
Night 16	70	2.94 (1.04)	18.9	68	3.11 (0.84)	22.5	0.157	68	2.92 (0.99)	18.5	0.747
Avg N15,16	71	3.08 (0.89)	21.7	69	3.29 (0.76)	26.8	0.051	70	3.04 (0.95)	20.9	0.841
Night 29	69	2.87 (1.18)	17.7	68	2.88 (1.06)	17.8	0.799	69	2.74 (1.11)	15.4	0.864
Night 30	67	3.00 (1.05)	20.2	67	3.07 (1.02)	21.6	0.699	68	2.74 (1.18)	15.5	0.322
Avg N29,30	69	3.03 (0.99)	20.8	69	3.08 (0.86)	21.9	0.557	69	2.83 (1.03)	17.0	0.487

Basing the analyses on the ITT-LOCF population instead of observed cases did not result in any changes to the statistical significance conclusions. Therefore, it seems relatively unlikely that the dropouts would have much impact on the results of the analyses of LPS if their data was complete.

Other Secondary Endpoints

The estimated correlation between Sleep Efficiency (SE) overall and Total Sleep Time (TST) is 1.00 because SE overall is defined as TST/TIB and TIB was almost always 480 minutes. So, SE all night seems to be redundant endpoint if considered after TST.

The distribution of Wake Time after Sleep (WTAS) is very asymmetric (skewed). Most values are 0 minutes, e.g., overall 76% were 0 on Night 1, but some extend far above 0. Therefore,

because of the highly non-normal distribution of WTAS the ANCOVA analysis used for the other endpoints is not very appropriate or reliable for WTAS. This reviewer investigated Wilcoxon rank sum tests comparing the sums of the ranks of the WTAS values between the groups as well as logistic regression models comparing the odds of the WTAS being greater than 0 between the groups. The results are shown in Table 21. The p-values from the logistic regression are not shown but they did not lead to any different significance conclusions. The sponsor reported that there was a significant difference between 6mg and placebo on night 1 based on ANCOVA analysis but overall 76% of the WTAS values were 0 which makes the normality assumption highly questionable. Furthermore, the Wilcoxon and logistic regression analysis methods which avoid this normality assumption did not result in nominal significance (Wilcoxon $p=0.0896$ and logistic $p=0.1811$). Either way no matter which analysis one considers most appropriate none of the approaches found a significant difference on Night 29, so even if there was an effect there is no compelling evidence that it lasted beyond Night 1.

Table 21 Study 501: Wake Time After Sleep by Night (Observed Cases)

Night	PLACEBO		3 MG		Difference from Placebo LSMEAN (S.E.)	Wilcoxon rank sum test vs. Placebo p-value	6 MG		Difference from Placebo LSMEAN (S.E.)	Wilcoxon rank sum test vs. Placebo p-value
	N	MEAN (S.D.)	N	MEAN (S.D.)			N	MEAN (S.D.)		
Baseline	76	5.6 (12.4)	76	8.4 (16.9)	2.6 (2.2)	0.5347	76	5.1 (9.1)	-0.9 (2.2)	0.7577
Night 1	72	6.4 (15.5)	75	0.7 (3.7)	-5.9 (1.5)	0.0039	73	1.1 (4.6)	-4.9 (1.5)	0.0896
Night 15	70	5.5 (20.1)	69	2.8 (8.7)	-3.4 (2.2)	0.4127	70	2.1 (7.8)	-3.2 (2.2)	0.7325
Night 29	69	5.8 (15.5)	68	3.2 (8.4)	-3.0 (2.0)	0.4906	69	2.7 (9.9)	-2.8 (1.9)	0.0967

There was no significant difference between Doxepin 6 mg and placebo on the objective number of awakenings after sleep onset (NAASO) at night 1 (6 mg vs. placebo $p=0.1378$) or night 29 ($p=0.1781$). This, together with the (b) (4) results, may raise questions about the sponsor's claim of an effect towards (b) (4) which they based on the sleep efficiency endpoint results.

Summary of Results

Table 22 summarizes the p-values from observed cases analyses for the endpoints considered most important for this review by the division of Neurology, for each of the two nights at each visit, as well as their average. These p-values are not adjusted for multiplicity which is a big consideration here. The Division's recommended approach was to start with the high dose at the latest time and work to the first time, first for oWASO, then for sWASO, LPS, and finally for LSO. If an intermediate test is not significant at 0.05 then no further testing of this sequence of tests should be done in order to control the experimentwise type I error at 0.05.

- Objective WASO was generally significantly reduced at each time for both 6 mg and 3 mg as compared to placebo.

- Subjective WASO was not significantly reduced on the first nights (sponsor designated as primary) of Visit 6 and Visit 5 for the 6 mg group vs. placebo. However, if we look at the 2nd night of each visit or the average of nights 1 and 2 then subjective WASO appears to be significantly reduced for the 6 mg group as compared to placebo.
- Objective LPS was only significant at Visit 4 and even that was not consistent over both nights. So, it is not clear that there is an effect on objective LPS beyond the first night of application of the drug. The same was true for the subjective LSO and since it is lower in the testing hierarchy no claims on LSO should be possible.

Table 22 Study 501: Results for 1st Night, 2nd Night, and Average of two at each Visit (OC)

Endpoint	Group	Night of Visit	p-value for comparison with placebo		
			Visit 6 (Night 29,30)	Visit 5 (Night 15,16)	Visit 4 (Night 1,2)
oWASO	6 mg	1 st	0.001	0.001	<0.001
		2 nd	<0.001	0.001	<0.001
		Avg	<0.001	<0.001	<0.001
	3 mg	1 st	0.017	0.003	<0.001
		2 nd	0.002	0.001	0.006
		Avg	0.001	<0.001	<0.001
sWASO	6 mg	1 st	0.628	0.202	<0.001
		2 nd	0.001	0.042	0.009
		Avg	0.020	0.042	<0.001
	3 mg	1 st	0.648	0.151	<0.001
		2 nd	0.059	0.248	0.257
		Avg	0.343	0.103	0.005
oLPS	6 mg	1 st	0.864	0.592	0.001
		2 nd	0.322	0.747	0.183
		Avg	0.487	0.841	0.007
	3 mg	1 st	0.800	0.227	0.006
		2 nd	0.699*	0.157*	0.282
		Avg	0.557*	0.051*	0.010
sLSO	6 mg	1 st	0.651*	0.145	0.049
		2 nd	0.763	0.452	0.809*
		Avg	0.699*	0.069	0.284
	3 mg	1 st	0.237	0.907*	0.126
		2 nd	0.518	0.649*	0.820
		Avg	0.334*	0.944*	0.187

*sign of t-statistic favors placebo

The placebo group had a modest within group increase in average subjective WASO (59.6 vs. 70.7, p=0.08) between Night 1 and Night 2 of the last visit, whereas, the 6 mg group had a modest within group decrease in average WASO (58.2 vs. 49.7, p=0.07) between Night 1 and Night 2 of the last visit (Visit 6 or the 3rd Post-Baseline Visit). This may help to explain the difference in significance of the 6 mg vs. placebo comparison between Night 1 and Night 2 of the last visit.

The placebo group had a nominally significant within group decrease in LPS (geometric mean 28.0 vs. 21.2, $p=0.0165$) between Night 1 and Night 2 of Visit 4 (first post-baseline visit). This may help to explain the difference in significance of the 6 mg vs. placebo comparison between Night 1 and Night 2 of Visit 4.

3.1.4 Study 502

The first subject was enrolled on 08 February 2006 and the last subject completed the study on 22 June 2006. All Investigators in this study were based in the United States.

The objectives of this study were as follows:

- To evaluate the effect of doxepin 6 mg relative to placebo on sleep onset in a model of transient insomnia.
- To assess the safety of doxepin 6 mg in healthy adult subjects.

The final protocol, protocol amendment 1, is dated 20 January 2006.

Efficacy Variables

The primary efficacy variable is LPS as determined by PSG assessment. The key secondary efficacy endpoint is WASO as determined by PSG assessment. Other secondary efficacy PSG endpoints include: TST, SE (whole night, by third of the night, and by hour), latency to Stage 2 sleep, WTDS, WTAS, NAASO, and sleep architecture (percentages and minutes of Stage 1, 2 and 3-4 nonREM [NREM] sleep, percentages and minutes of REM and NREM sleep, and latency to REM sleep). Subjective sleep assessments as noted on the morning questionnaires include: LSO, sTST, sNAASO, sTWT, sWASO, and sleep quality.

Analysis Methods for the Efficacy Analysis

Efficacy analyses were to be performed on the ITT analysis set.

The ITT analysis set was to include all randomized subjects who have a corresponding PSG efficacy assessment of LPS at Night 1. Data were to be analyzed as randomized.

Analysis of the Primary Efficacy Endpoint

The primary efficacy variable is LPS as determined by PSG assessment. The null hypothesis to be tested is that the LPS values for the two treatments are equal. The alternative hypothesis is that the LPS values are different for doxepin HCl 6 mg compared with placebo. Hypothesis tests for the comparison of doxepin HCl at 6 mg versus placebo were to be analyzed using an analysis of variance (ANOVA) model with main effects for treatment and center. If the data were inconsistent with the assumptions of ANOVA, an appropriate transformation of the data might be performed or appropriate non-parametric tests might be used to evaluate the differences among treatment groups.

Secondary Efficacy Endpoints

Continuous secondary efficacy variables (including both the key and other secondary endpoints) were to be analyzed using the same methods used to compare the LPS values, i.e., using an ANOVA model as appropriate.

Sleep quality was to be summarized using methods for continuous data. Differences between treatments were to be analyzed using an ANOVA model with main effects for treatment and center. For each outcome, if the overall frequency distribution (doxepin HCl 6 mg and placebo combined) indicated that subjects reported outcomes in fewer than five categories, differences between treatment groups may be analyzed using the Cochran-Mantel-Haenszel chi-square (row meanscore) test stratifying by center, assuming the categories of response are equally spaced.

Handling of Missing Individual Items of Efficacy Assessments

Before breaking the randomization blind, a complete review of all data was to be conducted in order to account for all missing values and protocol violations. Data that were established to be truly missing were to be documented with, if possible, the reason for the missing values. A missing value code(s) was to be documented and incorporated into the database. All analyses were to be based on observed data; missing data was not to be imputed.

Sample Size Determination

Approximately 500 subjects were to be randomized into the study. This sample size was based on results from a similar transient insomnia study conducted in 375 healthy adults. Results from this study showed that the estimated standard deviation of LPS was approximately 22 minutes in the placebo group, with smaller estimated standard deviations in the remaining two treated groups. Assuming a standard deviation of 22 minutes, 250 subjects per arm provides greater than 85% power to detect a difference in LPS between the doxepin HCl 6 mg and placebo group of at least 6 minutes, using a two-sided two-sample t-test at the 5% level of significance.

Changes introduced in the Statistical Analysis Plan (SAP) dated September 21, 2006.

A step-down procedure was to be used to control for multiple comparisons of the primary and key secondary endpoints. If the comparison of the doxepin 6 mg group to placebo with respect to LPS was statistically significant ($p \leq 0.05$), the comparison with respect to WASO was to be made. No other adjustments for multiple comparisons of the remaining efficacy endpoints were to be conducted.

3.1.4.1 Sponsor's Results

All 565 randomized subjects (282 subjects in the placebo group and 283 subjects in the doxepin 6 mg group) completed the study. A total of six study centers in the US randomized subjects into the study. Each center randomized 42 to 144 subjects, inclusive.

Demographic and other baseline characteristics are summarized by treatment group based on the safety analysis set in Table 23. Subjects were female (55%) and male (45%). Gender distribution was well-balanced between the treatment groups. The mean age of subjects was 35.5 years.

Subjects were White (50%), Hispanic (32%), Black/African American (15%), and Asian (1%), Native Hawaiian or Other Pacific Islander (<1%), and Other (1%).

Table 23 Study 502: Baseline Demographics

Variable	Placebo (N=282)	Doxepin 6 mg (N=283)	Total (N=565)
Age (years)			
Mean (SD)	35.9 (8.13)	35.2 (8.20)	35.5 (8.17)
Range	25–55	25–55	25–55
Sex [n (%)]			
Male	134 (48%)	123 (43%)	257 (45%)
Female	148 (52%)	160 (57%)	308 (55%)
Race/Ethnicity [n (%)]			
White	151 (54%)	133 (47%)	284 (50%)
Black/African American	39 (14%)	45 (16%)	84 (15%)
Hispanic	87 (31%)	93 (33%)	180 (32%)
Native Hawaiian or Other Pacific Islander	0 (0%)	1 (<1%)	1 (<1%)
Asian	2 (1%)	6 (2%)	8 (1%)
Other	3 (1%)	5 (2%)	8 (1%)
Weight (kg)			
Mean (SD)	73.4 (13.27)	73.9 (13.46)	73.7 (13.36)
Range	46–109	50–120	46–120
BMI (kg/m²)			
Mean (SD)	25.1 (2.90)	25.2 (2.75)	25.2 (2.82)
Range	19.2–32.1	20.0–32.3	19.2–32.3

Analysis of the Primary Efficacy Variable – Latency to Persistent Sleep

The primary efficacy variable, LPS, was measured using PSG recordings.

A summary of LPS by treatment group using the ITT analysis set is presented in Table 24. There was a statistically significant decrease ($p < 0.0001$) in the mean LPS for the doxepin 6 mg group compared with the placebo group. The LS mean estimate for LPS was 13.0 minutes shorter for the doxepin 6 mg group compared with the placebo group.

Table 24 Study 502: Primary Efficacy Variable – LPS on Night 1: ITT Analysis Set

LPS (minutes)	Placebo (N=282)	Doxepin 6 mg (N=283)
Subjects	282	282
Mean (SD)	33.6 (36.87)	20.6 (18.93)
Median (Range)	19.3 (0.0–236.0)	15.0 (0.5–165.0)
LS Mean (Std. Err.)	32.9 (1.83)	20.0 (1.83)
Difference of LS Mean (Std. Err.)		-13.0 (2.44)
95% CI of LS Mean Difference		(-17.8, -8.2)
p-value ¹		$p < 0.0001$

¹ p-value for comparing treatments was determined from an ANOVA model that included main effects for treatment and center.

Secondary Endpoints

A summary of WASO by treatment group based on the ITT analysis set is provided in Table 25. There was a statistically significant decrease in the mean WASO for the doxepin 6 mg group compared with the placebo group. The LS mean WASO estimate was 39.1 minutes shorter for the doxepin 6 mg group compared with the placebo group.

Table 25 Study 502: WASO on Night 1: ITT Analysis Set

WASO (minutes)	Placebo (N=282)	Doxepin 6 mg (N=283)
Subjects	n=281	n=281
Mean (SD)	77.5 (62.11)	38.4 (31.70)
Median (Range)	60.5 (6.5–364.0)	28.5 (1.0–189.0)
LS Mean (Std. Err.)	79.4 (3.11)	40.4 (3.11)
Difference of LS Mean (Std. Err.)		-39.1 (4.16)
95% CI of LS Mean Difference		(-47.2, -30.9)
p-value ¹		p<0.0001

¹ p-value for comparing treatments was determined from an ANOVA model that included main effects for treatment and center.

There was a nominally statistically significant improvement in the mean all night Sleep Efficiency (SE) for the doxepin 6 mg group compared with the placebo group. The LS mean SE was greater (improved) for the doxepin 6 mg group by 10.6% compared with the placebo group.

Table 26 Study 502: SE Overall and by First, Second, and Final Third of the Night on Night 1: ITT Analysis Set

SE Variable	Placebo (N=282)	Doxepin 6 mg (N=283)
SE–Overall (%)	n=281	n=281
Mean (SD)	77.9 (14.47)	88.6 (8.32)
Median (Range)	80.6 (18.2–98.3)	91.0 (35.0–99.3)
LS Mean (Std. Err.)	77.6 (0.75)	88.3 (0.75)
Difference of LS Mean (Std. Err.)		10.6 (0.99)
95% CI of LS Mean Difference		(8.7, 12.6)
p-value ¹		p<0.0001
SE–First Third of the Night (%)	n=281	n=281
Mean (SD)	69.6 (22.71)	82.6 (13.79)
Median (Range)	77.2 (0.0–98.4)	87.2 (6.3–99.4)
LS Mean (Std. Err.)	69.5 (1.17)	82.5 (1.17)
Difference of LS Mean (Std. Err.)		13.0 (1.56)
95% CI of LS Mean Difference		(9.9, 16.1)
p-value ¹		p<0.0001
SE–Second Third of the Night (%)	n=281	n=281
Mean (SD)	82.5 (20.13)	91.9 (9.67)
Median (Range)	90.9 (0.0–99.4)	94.7 (10.3–100.0)
LS Mean (Std. Err.)	81.8 (0.99)	91.2 (0.99)
Difference of LS Mean (Std. Err.)		9.4 (1.33)
95% CI of LS Mean Difference		(6.8, 12.0)
p-value ¹		p<0.0001
SE–Final Third of the Night (%)	n=281	n=281
Mean (SD)	81.7 (22.02)	91.2 (9.48)
Median (Range)	91.9 (1.6–100.0)	94.1 (29.4–100.0)
LS Mean (Std. Err.)	81.6 (1.07)	91.1 (1.07)
Difference of LS Mean (Std. Err.)		9.5 (1.43)
95% CI of LS Mean Difference		(6.7, 12.3)
p-value ¹		p<0.0001

¹ p-value for comparing treatments was determined from an ANOVA model that included main effects for treatment and center.

3.1.4.2 Reviewer's Comments

This reviewer verified the sponsor's primary analysis result which showed that the Doxepin 6 mg group had a statistically significantly shorter average Latency to Persistent Sleep ($p < 0.0001$). As this was only a 1 night study there were no missing data. This reviewer also verified the sponsor's result for WASO ($p < 0.0001$). Note that no baseline measures of LPS or WASO were recorded. Therefore, we can not be certain that the groups were completely balanced with respect to these measures at baseline.

Overall, eighty four percent of the WTAS values on night 1 were 0 minutes. A Wilcoxon rank sum test ($p = 0.01$) and a logistic regression ($p = 0.022$), methods which avoid the questionable normality assumption for WTAS, suggested that the 6 mg group was nominally significantly better than placebo in terms of WTAS. However, WTAS was not specified as a key secondary endpoint or adjusted for multiple comparisons.

There was no significant difference between Doxepin 6 mg and placebo on the objective number of awakenings after sleep onset (NAASO) for Treatment Night 1 ($p = 0.2252$). This may raise questions about the sponsor's claim of an effect towards (b) (4) based on the results for the sleep efficiency endpoint.

3.1.5 Study 503

Title: A Phase III, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Multicenter Study to Assess the Long Term Efficacy and Safety of Doxepin Hcl in Primary Elderly Insomnia Patients with Sleep Maintenance Difficulties

The first subject was enrolled on 06 September 2005 (randomization date) and the last subject completed on 28 September 2006. All Investigators in this study were based in the United States.

3.1.5.1 Study Design and Statistical Methods

Primary Objective: To evaluate the sleep efficacy of two dose levels of doxepin HCl (doxepin) relative to placebo in elderly subjects with sleep maintenance difficulties due to primary insomnia.

Secondary Objectives: (1) To evaluate the long-term efficacy of two dose levels of doxepin in elderly subjects with primary insomnia, and (2) to evaluate the safety of doxepin in elderly subjects with primary insomnia when administered for 12 weeks.

The final version of the protocol, Protocol Amendment 2, was dated July 28, 2006:
The primary efficacy variable is WASO at Visit 3 (Day 1).

The primary alternative hypothesis is that the WASO values at Visit 3 (Day 1) are different for doxepin HCl 3 mg compared to placebo; the comparison of doxepin HCl 1 mg to placebo is secondary. Hypothesis tests for the comparison of doxepin HCl at 1 mg and 3 mg versus placebo

will be analyzed using an analysis of covariance (ANCOVA) model that includes the main effects for treatment and center with the baseline WASO as a covariate. Baseline WASO is defined as the mean of PSG Screening Nights -6 and -5.

If the data are inconsistent with the assumptions of ANCOVA, an appropriate transformation of the data may be performed or appropriate non-parametric tests will be used to evaluate the differences among treatment groups.

The following changes/additions were described in the statistical analysis plan which was dated November 29, 2006, which the sponsor asserted was before unblinding.

Analysis Population and Missing Data Handling

The primary efficacy analysis was to be based on the ITT analysis set using observed data only; missing data was not to be imputed. Additional sensitivity analyses that impute missing data were also to be performed for the primary and key secondary variables. The first sensitivity analysis was to impute missing data using the last observation carried forward (LOCF) method. The second sensitivity analysis was to impute missing data using the baseline observation carried forward (BOCF) method.

Multiplicity Adjustments

The primary endpoint, WASO at Visit 3, was to be assessed at the 5% level of significance (two-sided) using a linear contrast within the context of the ANCOVA model to compare the mean WASO for the doxepin 3 mg group to placebo. If statistically significant, the duration of the effect of doxepin on WASO across the entire study period was to be assessed by comparing the mean WASO for doxepin 3 mg to that for placebo at Visit 5 (Day 29) and Visit 7 (Day 85). If all three of these comparisons were statistically significant, the difference between doxepin 3 mg and placebo at Visit 3 was to be assessed for the key secondary endpoints of sTST (from the morning questionnaire), LPS, and SE in Hour 8. Each comparison was to be assessed only if the preceding comparison achieved statistical significance. According to the sponsor this sequential testing procedure controls the overall Type 1 error rate for these comparisons at the 5% level of significance (Westfall et al, 1999).

Specifically, the order of testing for statistical significance will be as follows:

- WASO at Visit 3
- WASO at Visit 5
- WASO at Visit 7
- sTST at Visit 3
- LPS at Visit 3
- SE in Hour 8 at Visit 3

The comparison of doxepin at 1 mg to placebo will follow the same sequential testing procedure, but only for those parameters that achieved statistical significance for the comparison of doxepin 3 mg to placebo.

No other adjustments for multiplicity will be made.

Reviewer's Note: The sponsor's multiplicity adjustment approach does not strongly control the experimentwise type I error. For example, suppose that the true means for the WASO at Visit 3

are markedly different for 3 mg and placebo, but they are the same for 3 mg and placebo at Visit 5 and the same for 1 mg and placebo at visits 3 and 5. Because the null hypothesis for 3 mg vs. placebo at visit 3 will be rejected (with probability essentially = 1 if the true difference is large enough), the two null hypotheses for 3 mg vs. placebo at Visit 5 and 1 mg vs. placebo at Visit 3 will always be carried out.

Therefore, the probability of at least one type I error is:

$$\begin{aligned} & \{Pr(\text{null for 3 mg vs. placebo at visit 5 rejected AND null for 1 mg vs. placebo at visit 3 Not rejected})+ \\ & Pr(\text{null for 3 mg vs. placebo at visit 5 Not rejected AND null for 1 mg vs. placebo at visit 3 rejected})+ \\ & Pr(\text{Nulls for both 3 mg vs. placebo at Visit 5 and 1 mg vs. placebo at Visit 3 are rejected})\} \\ & =.05*.95 + .05*.95+ .05*.05=.0975 > 0.05 \end{aligned}$$

3.1.5.2 Patient Disposition

A summary of disposition for all randomized subjects is provided in Table 27. All 240 randomized subjects (100%) received double-blind study drug and were included in the Safety Analysis Set and the ITT Analysis Set. Overall, 214 subjects (89%) completed the study. The 26 subjects (11%) who did not complete the study all discontinued during the Double-blind Treatment Period. Seven (3%) randomized subjects withdrew from the study due to an AE. An additional subject (Subject 32-7307) had an AE with an erroneous CRF entry for Action Taken of study drug discontinuation. (Study drug was interrupted due to an SAE and resumed; subsequently, the subject was withdrawn from the study due to a protocol violation.) There were no deaths reported during the study or within 30 days following administration of the last dose of study drug.

Table 27 Study 503: Patient Disposition

	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)	Total (N=240)
Completed the Study	70 (86%)	70 (91%)	74 (90%)	214 (89%)
Withdrew from the Study	11 (14%)	7 (9%)	8 (10%)	26 (11%)
Primary Reason for Discontinuation¹				
Adverse Event ²	3 (4%)	1 (1%)	3 (4%)	7 (3%)
Protocol Violation	2 (2%)	2 (3%)	1 (1%)	5 (2%)
Noncompliance	0 (0%)	3 (4%)	0 (0%)	3 (1%)
Consent Withdrawn	6 (7%)	0 (0%)	2 (2%)	8 (3%)
Lost to Follow-up	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Death	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other	0 (0%)	1 (1%)	2 (2%)	3 (1%)
Subjects with an AE Resulting in Discontinuation of Study Drug³	3 (4%)	2 (3%)	3 (4%)	8 (3%)

Note: No randomized subject withdrew from the study before receiving double-blind study drug.

¹ Primary reason for discontinuation as recorded on the Study Termination CRF page.

² Includes [Subject 78-7188](#) (doxepin 3 mg), whose primary reason for discontinuation was recorded as "Other: SAE."

³ Includes [Subject 32-7307](#) (doxepin 1 mg), who had an AE (pneumonia) with an erroneous CRF entry for Action Taken to Study Drug = Discontinued. This subject was withdrawn from the study due to a protocol violation.

SOURCE: SPONSOR STUDY REPORT PAGE 72

Approximately 10% of subjects in each doxepin treatment group withdrew from the study, compared with 14% of subjects in the placebo group. There were no important differences across treatment groups regarding the reasons for study withdrawal. The most frequent primary reasons for discontinuation were consent withdrawn (7%) and AEs (4%) for subjects in the placebo group; noncompliance (4%) and protocol violation (3%) for subjects in the doxepin 1 mg group; and AEs (4%) for subjects in the doxepin 3 mg group. Of the subjects withdrawn from the study due to noncompliance, Subject 08-7112, who was included in all three analysis sets, was noncompliant with regard to study medication. Another subject was participating in this study at two study centers, as Subject 06-7076 (doxepin 1 mg) and Subject 19-7020 (doxepin 1 mg); she was excluded from the PP Analysis Set. Data obtained during her participation at both study centers were included in the safety and ITT analysis sets. Thus, although Table 27 indicates three subjects withdrew due to noncompliance, only two subjects actually did so since one subject participated twice and had two subject numbers.

Subject 01-7303 in the placebo group was discontinued due to a protocol violation after receiving approximately 57 days of double-blind study medication once it was discovered that she had previously completed the study as Subject 06-7225 (also in the placebo group) at another study center. Subject 01-7303 was excluded from the PP Analyses Set. Data obtained during her participation at both study centers were included in the safety and ITT analysis sets.

3.1.5.2.1 Baseline Demographics

Demographic and other baseline characteristics are summarized by treatment group using the Safety Analysis Set in Table 28. Demographic and other baseline characteristics were similar across treatment groups. The study population was female (65%) and male (35%). The mean age was 71.4 years. Subjects were White (80%), Black/African American (9%), Hispanic (9%), Asian (1%), and Other (1%).

Table 28 Study 503: Baseline Demographics

Variable/Category	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)	Total (N=240)
Age (years)				
Mean (SD)	71.5 (5.50)	71.3 (5.23)	71.4 (4.88)	71.4 (5.19)
Range	65–93	64–85	65–88	64–93
Sex [n (%)]				
Male	33 (41%)	27 (35%)	25 (30%)	85 (35%)
Female	48 (59%)	50 (65%)	57 (70%)	155 (65%)
Race/Ethnicity [n (%)]				
White	67 (83%)	63 (82%)	63 (77%)	193 (80%)
Black/African American	6 (7%)	5 (6%)	10 (12%)	21 (9%)
Hispanic	4 (5%)	8 (10%)	9 (11%)	21 (9%)
Asian	1 (1%)	1 (1%)	0 (0%)	2 (1%)
Other	3 (4%)	0 (0%)	0 (0%)	3 (1%)
Weight (kg)				
Mean (SD)	78.8 (15.53)	75.5 (16.02)	75.5 (13.23)	76.6 (14.97)
Range	52–118	45–108	50–107	45–118
BMI (kg/m³)				
Mean (SD)	28.0 (4.77)	27.5 (5.42)	27.1 (4.37)	27.5 (4.85)
Range	20.0–44.1	18.3–41.8	18.8–39.7	18.3–44.1

3.1.5.3 Sponsor's Results

Thirty-one of the 35 study centers randomized subjects into the study. Data from 13 centers (Nos. 5, 10, 13, 16, 21, 27, 29, 35, 37, 42, 43, 45, and 68) with low enrollment (fewer than five subjects in the ITT Analysis Set) were pooled to form one pseudo-center, as described in the SAP.

Analysis of the Primary Efficacy Variable – WASO on Night 1

The primary efficacy variable was WASO on Night 1.

A summary of WASO at baseline and on Night 1 using the ITT Analysis Set is provided in Table 29. At baseline, the mean WASO was similar across the treatment groups. On Night 1, there were statistically significant decreases in the mean WASO for each

doxepin group, 1 mg and 3 mg, compared with placebo. The LS mean WASO estimate was shorter for the doxepin 1 mg and 3 mg groups by 17.8 minutes and 33.8 minutes, respectively, compared with the placebo group.

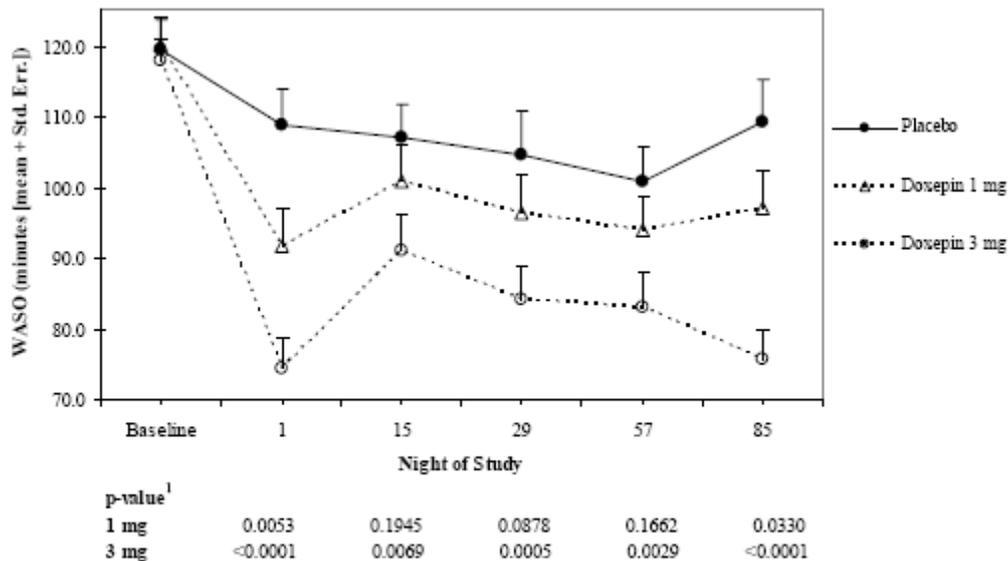
Table 29 Study 503: Objective WASO on Night 1 in ITT population

WASO (minutes)	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)
Baseline (Mean of Nights -6 and -5)			
Mean (SD)	119.5 (37.67)	120.1 (34.97)	117.9 (28.15)
Median (Range)	113.5 (62.5–294.5)	112.5 (65.3–204.8)	113.9 (74.8–209.5)
Night 1 (Visit 3)			
Mean (SD)	108.9 (46.01)	91.8 (47.09)	74.5 (37.88)
Median (Range)	107.0 (31.5–244.5)	77.5 (28.5–227.5)	67.8 (15.0–164.5)
Diff. of LS Mean (Std. Err.)		-17.8 (6.32)	-33.8 (6.23)
95% CI of LS Mean Diff.		(-30.3, -5.4)	(-46.0, -21.5)
p-value ¹		p=0.0053	p<0.0001

¹ p-value comparing each active treatment to placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using a linear contrast.

The sponsor’s analyses of the objective WASO over time are displayed in Figure 2. Note that the p-values are not adjusted for multiple comparisons.

Figure 2 Study 503: Objective WASO over Time



Note: Baseline is the mean of Nights -6 and -5.

¹ p-value comparing each active treatment versus placebo was obtained from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using a linear contrast.

Source: sponsor study report page 82

Subjective Total Sleep Time

Nominally significant improvements in sTST compared with placebo were observed in the doxepin 1 mg group at Day 86 and in the doxepin 3 mg group at Day 30 and the average of Days 2, 16, 30, 58, and 86. However, there were no significant differences between 3 mg and placebo or 1 mg and placebo at Visit 3, the time specified as first in the testing order by the sponsor. Therefore, the sponsor's prespecified multiplicity adjustment method only permits a claim on the objective WASO (at days 1, 15, 29, 57, and 85 for Doxepin 6 mg).

Table 30 Study 503: Subjective TST at Baseline, Day 2: ITT Analysis Set

sTST (minutes)	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)
Baseline (Mean of Days -5 and -4)	n=81	n=77	n=82
Mean (SD)	306.1 (67.43)	302.1 (71.91)	321.3 (60.59)
Day 2 (Visit 3)	n=78	n=77	n=82
Mean (SD)	322.0 (78.12)	315.1 (90.47)	353.1 (80.64)
p-value ¹		p=0.6905	p=0.0865

Source: study report page 112

Latency to Persistent Sleep

Note that the sponsor's prespecified multiplicity adjustment approach does not permit testing Latency to Persistent Sleep or at least not making any claims of significance for LPS. Latency to Persistent Sleep (LPS) at baseline, Night 1, Night 15, Night 29, Night 57, Night 85, and the average of Nights 1, 15, 29, 57, and 85 using the Observed Cases data in the ITT Analysis Set is summarized in Table 31.

At baseline, the geometric mean LPS was moderately lower in the doxepin 3 mg group. Statistically significant decreases in LPS were not observed at any timepoint in either doxepin group compared with placebo.

Latency to Persistent sleep for the average of Nights 1, 15, 29, 57, and 85 was not statistically significant compared with placebo for either doxepin group.

Table 31 Study 503: LPS over Time in ITT population (Sponsor's Results)

LPS (minutes)	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)
Baseline (Mean of Nights -6 and -5)	n=81	n=77	n=82
Mean (SD)	49.0 (27.34)	45.4 (25.25)	41.9 (22.65)
Median (Range)	47.0 (12.8–149.0)	38.0 (13.5–117.5)	38.3 (11.3–107.8)
Geometric Mean (SD)	41.9	39.3	35.8
Night 1 (Visit 3)	n=81	n=77	n=82
Mean (SD)	39.6 (29.28)	38.8 (29.58)	28.6 (20.53)
Median (Range)	31.0 (3.0–119.5)	29.5 (4.0–175.5)	22.8 (3.5–136.0)
Geometric LS Mean ¹	27.4	29.1	23.1
p-value ²		p=0.5733	p=0.1079
Night 15 (Visit 4)	n=78	n=74	n=80
Mean (SD)	45.1 (39.60)	44.0 (53.24)	40.2 (37.86)
Median (Range)	31.5 (1.5–175.5)	30.3 (7.0–364.5)	29.5 (4.0–225.5)
Geometric LS Mean ¹	30.2	31.1	30.9
p-value ²		p=0.8046	p=0.8388
Night 29 (Visit 5)	n=75	n=74	n=77
Mean (SD)	39.1 (42.38)	49.2 (51.23)	39.6 (40.01)
Median (Range)	26.5 (1.5–293.5)	31.8 (3.5–231.0)	24.5 (1.5–245.0)
Geometric LS Mean ¹	24.1	30.4	26.4
p-value ²		p=0.1268	p=0.5422
Night 57 (Visit 6)	n=71	n=71	n=75
Mean (SD)	30.8 (21.61)	37.0 (28.53)	35.2 (24.68)
Median (Range)	25.0 (2.5–96.5)	32.0 (2.0–122.5)	29.5 (1.5–139.0)
Geometric LS Mean ¹	22.4	26.7	28.9
p-value ²		p=0.1870	p=0.0522
Night 85 (Visit 7)	n=70	n=69	n=74
Mean (SD)	34.9 (32.96)	29.0 (26.45)	37.5 (32.74)
Median (Range)	25.3 (1.5–145.5)	22.5 (1.5–127.0)	26.3 (3.5–182.0)
Geometric LS Mean ¹	20.9	19.6	29.0
p-value ²		p=0.6493	p=0.0286
Average of Nights 1, 15, 29, 57 & 85	n=81	n=77	n=82
Mean (SD)	38.9 (25.83)	40.6 (27.46)	36.2 (24.13)
Median (Range)	31.6 (3.6–125.8)	30.3 (5.5–137.0)	31.7 (8.8–185.5)
Geometric LS Mean ¹	29.9	32.9	32.1
p-value ²		p=0.2688	p=0.4032

¹ Analysis was performed on log-transformed data. If the value was zero, log(0.25 minutes) was used. Geometric means were converted to original scale by taking the anti-log.

² p-value comparing each active treatment versus placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using a linear contrast.

Source: page 100 of sponsor's study report

The sponsor's results for Sleep Efficiency in the 8th hour of the night are shown in Table 32. They found that both Doxepin group's average differences from placebo were nominally significant on Night 1.

Table 32 Study 503: SE in Hour 8 at Baseline, Night 1: ITT Analysis Set

Sleep Efficiency (%) in Hour 8	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)
Baseline (Mean of Nights -6 and -5)	n=81	n=77	n=82
Mean (SD)	62.8 (21.00)	59.3 (25.47)	60.4 (19.69)
Median (Range)	65.8 (5.4–99.2)	66.3 (1.3–98.3)	60.2 (9.6–96.7)
Night 1 (Visit 3)	n=81	n=77	n=82
Mean (SD)	60.0 (31.10)	68.4 (26.77)	76.5 (20.51)
Median (Range)	66.7 (0.0–99.2)	75.8 (0.0–100.0)	82.9 (0.0–100.0)
Diff. of LS Mean (Std. Err.)		9.6 (4.12)	17.4 (4.05)
95% CI of LS Mean Diff.		(1.4, 17.7)	(9.4, 25.3)
p-value ¹		p=0.0211	p<0.0001

Source: Study report page 96

Number of Awakenings After Sleep Onset

Number of Awakenings After Sleep Onset at baseline and Night 1 using the ITT Analysis Set are summarized in Table 33. At baseline, the mean NAASO was similar across the treatment groups. Statistically significant improvements in the mean NAASO were not observed at any timepoint in either doxepin group compared with placebo.

Table 33 Study 503: Number of Awakenings after Sleep Onset in ITT population

NAASO	Placebo (N=81)	Doxepin 1 mg (N=77)	Doxepin 3 mg (N=82)
Baseline (Mean of Nights -6 and -5)	n=81	n=77	n=82
Mean (SD)	13.6 (4.75)	14.4 (4.62)	13.3 (4.26)
Median (Range)	13.5 (4.0–26.0)	13.5 (6.5–27.5)	13.0 (5.5–25.5)
Night 1 (Visit 3)	n=81	n=77	n=82
Mean (SD)	13.2 (5.46)	14.3 (6.44)	14.0 (6.19)
Median (Range)	11.0 (3.0–28.0)	13.0 (3.0–36.0)	14.0 (3.0–34.0)
p-value ¹		p=0.6109	p=0.2930

¹p-value comparing each active treatment versus placebo was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate using a linear contrast.

3.1.5.4 Reviewer’s Comments

Recall from section 3.1.5.1 that the sponsor prespecified the following testing hierarchy in an attempt to control the experimentwise type I error at 0.05:

- WASO at Visit 3
- WASO at Visit 5
- WASO at Visit 7
- sTST at Visit 3
- LPS at Visit 3
- SE in Hour 8 at Visit 3

There was no missing data for the primary analysis of objective WASO at Visit 3.

This reviewer verified the sponsor’s primary analyses of the objective WASO. This was based on an observed cases analysis, i.e., with no imputation for missing data. The result was still

significant if missing data were imputed using either a) BOCF (baseline carried forward) or b) LOCF(last observation carried forward). The results of various imputations for missing data are summarized in Table 34.

The 1 mg group average difference from placebo in terms of the objective WASO did not reach nominal significance at nights 57, 29, or 15 but did at nights 85 and 1. The statistical significance conclusions for 1 mg were unchanged by LOCF or BOCF imputations for missing data.

Table 34 Study 503: Comparison of Results from various Imputation Methods for Missing objective WASO data

Pop	Night	PLACEBO		1 MG				3 MG			
		N	MEAN (S.D.)	N	MEAN (S.D.)	LS MEAN (S.E.) Difference from placebo	p-value	N	MEAN (S.D.)	LS MEAN (S.E.) Difference from placebo	p-value
OC	Baseline	81	119.5 (37.7)	77	120.1 (35.0)	0.9 (5.4)	0.867	82	117.9 (28.1)	-1.2 (5.3)	0.821
	1	81	108.9 (46.0)	77	91.8 (47.1)	-17.8 (6.3)	0.0053	82	74.5 (37.9)	-33.8 (6.2)	<0.0001
	15	78	107.1 (41.1)	74	100.8 (46.2)	-8.5 (6.5)	0.1945	80	91.0 (46.7)	-17.4 (6.4)	0.0069
	29	75	104.6 (53.5)	74	96.4 (45.3)	-10.8 (6.3)	0.0878	77	84.3 (40.9)	-22.1 (6.3)	0.0005
	57	71	100.8 (42.1)	71	94.0 (38.7)	-8.5 (6.1)	0.1662	75	83.0 (42.5)	-18.3 (6.0)	0.0029
	85	70	109.2 (50.8)	69	97.0 (44.2)	-14.6 (6.8)	0.0330	74	75.7 (37.5)	-33.2 (6.7)	<0.0001
LOCF	15	81	107.5 (42.2)	77	99.9 (46.0)	-8.0 (6.4)	0.209	82	90.2 (46.6)	-16.8 (6.3)	0.008
	29	81	105.0 (53.4)	77	95.7 (45.1)	-10.0 (6.4)	0.121	82	87.5 (45.1)	-16.9 (6.3)	0.008
	57	81	101.4 (42.4)	77	93.1 (39.5)	-8.5 (6.0)	0.155	82	86.1 (46.3)	-14.5 (5.9)	0.015
	85	81	107.4 (50.0)	77	95.0 (44.3)	-13.0 (6.5)	0.046	82	79.7 (42.5)	-26.5 (6.4)	<0.001
BOCF	15	81	109.5 (45.7)	77	101.4 (45.9)	-8.8 (6.3)	0.167	82	91.5 (46.3)	-17.5 (6.2)	0.005
	29	81	107.3 (56.0)	77	97.2 (45.1)	-11.1 (6.1)	0.071	82	86.4 (40.8)	-20.5 (6.0)	0.001
	57	81	105.6 (46.3)	77	95.6 (39.0)	-10.4 (5.8)	0.073	82	85.2 (41.9)	-19.8 (5.7)	0.001
	85	81	112.6 (52.2)	77	97.9 (43.5)	-15.3 (6.3)	0.015	82	78.7 (37.5)	-32.9 (6.2)	<0.001

Another approach to assessing the impact of missing data on the results is a mixed model for repeated measures. The analysis of this model agreed with the other three methods considered here. Therefore, it seems relatively unlikely that the dropouts (11 placebo, 8 Doxepin 1 mg, and 8 Doxepin 3 mg) would alter the results if their data was complete.

Objective Latency to Persistent Sleep

Note that at Night 85 (the final visit) the 3 mg group was nominally significantly worse than the placebo group in terms of the LPS (p=0.0286). At night 85, 14% of placebo and 10% of 3 mg had dropped out of the study. However, the same conclusion was also reached using either

LOCF or BOCF imputations for missing data. There were no nominally significant differences favoring 3 mg over placebo in terms of LPS.

Table 35 Study 503: Objective Latency to Persistent Sleep by Night for OC and ITT-LOCF analyses

POP-ULATION	TIME	PLACEBO			1 MG				3MG			
		N	MEAN of Log LSO (S.D.)	GEO-METRIC MEAN of LSO	N	MEAN of Log LSO (S.D.)	GEO-METRIC MEAN of LSO	p-value for diff from placebo	N	MEAN of Log LSO (S.D.)	GEO-METRIC MEAN of LSO	p-value for diff from placebo
OC	Night 1	81	3.39 (0.82)	29.7	77	3.42 (0.70)	30.6	0.5733	82	3.14 (0.67)	23.1	0.1079
	Night 15	78	3.43 (0.94)	30.9	74	3.43 (0.78)	30.8	0.8046	80	3.35 (0.84)	28.6	0.8388
	Night 30	75	3.24 (1.00)	25.4	74	3.46 (0.95)	31.7	0.1268	77	3.26 (0.96)	26.0	0.5422
	Night 57	71	3.14 (0.82)	23.2	71	3.28 (0.89)	26.7	0.1870	75	3.32 (0.76)	27.7	0.0522
	Night 85	70	3.10 (1.03)	22.2	69	3.01 (0.92)	20.4	0.6493	74	3.31 (0.81)	27.4	0.0286
LOCF	Night 15	81	3.42 (0.96)	30.6	77	3.42 (0.80)	30.6	0.7847	82	3.35 (0.83)	28.5	0.7795
	Night 30	81	3.26 (1.00)	26.0	77	3.45 (0.96)	31.4	0.1345	82	3.28 (0.96)	26.6	0.4453
	Night 57	81	3.20 (0.84)	24.6	77	3.34 (0.94)	28.1	0.2116	82	3.35 (0.76)	28.4	0.0822
	Night 85	81	3.15 (1.02)	23.4	77	3.06 (0.98)	21.4	0.7092	82	3.34 (0.81)	28.2	0.0386

Subjective WASO

The difference between 3 mg and placebo on Subjective WASO was significant at night 85 ($p=0.0153$) but not at nights 57, 15, or 1. Also, the 3 mg vs. placebo difference on the subjective total sleep time, which the sponsor prefers over subjective WASO and which they specified as key secondary, was not significant at night 85, 57, 15, or 1.

Table 36 summarizes the analyses of subjective WASO by night for observed cases as well as for ITT with LOCF imputation for dropouts that had some post-baseline subjective WASO data. There was just one case in which the significance conclusion was different for observed cases and LOCF analyses (Night 29, 6 mg vs. placebo: OC $p=0.0296$; LOCF $p=0.0705$). Therefore, for the most part, it appears relatively unlikely that the dropouts missing data would impact the results of the observed cases analyses if it was available.

Table 36 Study 503: Comparison of Observed Case and ITT-LOCF results for subjective WASO by Night

	Night	PLACEBO		1 MG				3MG			
		N	MEAN (S.D.)	N	MEAN (S.D.)	LS MEAN (S.E.) Difference from placebo	p-value	N	MEAN (S.D.)	LS MEAN (S.E.) Difference from placebo	p-value
OC	Night 1	78	89.0 (66.0)	74	84.4 (80.6)	-1.9 (10.2)	0.8497	82	62.1 (62.7)	-19.2 (10.0)	0.0561
	Night 15	76	87.8 (61.4)	72	86.7 (61.8)	-1.7 (9.4)	0.8571	78	67.9 (64.7)	-16.7 (9.3)	0.0729
	Night 29	75	92.6 (74.6)	73	73.5 (68.2)	-19.4 (10.0)	0.0531	77	68.0 (49.3)	-21.7 (9.9)	0.0296
	Night 57	69	78.0 (64.2)	70	78.1 (75.9)	3.4 (10.2)	0.7417	74	62.5 (47.9)	-5.9 (10.1)	0.5627
	Night 85	69	88.6 (62.9)	69	57.9 (61.1)	-29.7 (10.1)	0.0037	74	56.7 (64.3)	-24.5 (10.0)	0.0153
LOCF	Night 15	80	88.5 (60.6)	77	84.8 (60.8)	-0.7 (9.2)	0.9360	82	68.2 (63.8)	-14.8 (9.1)	0.1051
	Night 29	81	91.7 (72.8)	77	71.5 (67.2)	-17.1 (9.6)	0.0763	82	69.2 (49.7)	-17.2 (9.5)	0.0705
	Night 57	81	81.4 (71.8)	77	77.1 (74.5)	-0.9 (10.1)	0.9284	82	65.0 (48.5)	-10.4 (10.0)	0.3005
	Night 85	81	86.2 (60.5)	77	59.6 (60.4)	-23.5 (9.4)	0.0128	82	59.4 (63.5)	-21.4 (9.3)	0.0215

Note: A few patients had their first sWASO measurement at night 16 which explains how the sample size for LOCF at night 16 can be greater than for OC at Night 1

Subjective Total Sleep Time

Subjective Total Sleep Time (sTST) was specified as the key secondary endpoint in the hierarchy of secondary endpoints. The difference between 3 mg and placebo on sTST was not statistically significant at the first timepoint, Night 1. Therefore, according to the sponsor’s prespecified testing hierarchy, testing cannot proceed to the LPS or sleep efficiency endpoints. Even if LPS could be tested, since the comparison on LPS was not statistically significant at Night 1 and it was before Sleep Efficiency in the hierarchy, again, the Sleep Efficiency results can not be tested or considered statistically significant after adjusting for multiple comparisons.

Summary of Results and Multiplicity Issues for 1 mg

Strictly speaking, under the prespecified multiplicity adjustment method, in order to strongly control the experimentwise type I error at 0.05 the 1 mg dose hypotheses should not be tested

because there were hypotheses involving the 3 mg dose that were higher in the hierarchy that were not significant. If we were to do so anyways for exploratory purposes, we would find that testing would stop after the second test, objective WASO at Night 29, because this comparison with placebo was not significant at the 0.05 level. Therefore, the multiplicity adjustment plan would not permit testing sTST, LPS, or Sleep Efficiency hypotheses for 1 mg. Once again, we could not even test WASO for 1 mg if we strictly observe that all 3 mg vs. placebo hypotheses were to be tested first and not all were significant.

Other Secondary Endpoints

The distribution of Wake Time after Sleep (WTAS) is very asymmetric (skewed). Most values are 0 minutes, e.g., overall 60% were 0 on Night 1, but some extend far above 0. Therefore, because of the highly non-normal distribution of WTAS the ANCOVA analysis used for the other endpoints is not very appropriate or reliable for WTAS. This reviewer investigated Wilcoxon rank sum tests comparing the sums of the ranks of the WTAS values between the groups as well as logistic regression models comparing the odds of the WTAS being greater than 0 between the groups. The results are shown in Table 21. The p-values from the logistic regression are not shown but they were close to the Wilcoxon p-values and did not lead to any different significance conclusions.

Table 37 Study 503: Wake Time After Sleep by Night (OC)

NIGHT	PLACEBO		1 MG				3 MG			
	N	MEAN (S.D.)	N	MEAN (S.D.)	Difference from Placebo LSMEAN (S.E.)	Wilcoxon rank sum test vs. Placebo p-value	N	MEAN (S.D.)	Difference from Placebo LSMEAN (S.E.)	Wilcoxon rank sum test vs. Placebo p-value
Baseline	81	10.1 (15.0)	77	9.7 (16.3)	0.1 (2.5)	0.5108	82	9.8 (15.3)	0.3 (2.5)	0.8795
NIGHT 1	81	10.7 (22.6)	77	8.9 (18.8)	-1.4 (3.0)	0.7082	82	4.7 (14.6)	-5.7 (2.9)	0.1405
NIGHT 15	78	14.1 (26.6)	74	13.2 (21.2)	-0.4 (3.7)	0.2818	80	8.6 (21.3)	-4.9 (3.6)	0.2782
NIGHT 29	75	13.3 (24.0)	74	11.0 (19.6)	-2.1 (3.2)	0.7503	77	6.7 (19.9)	-6.1 (3.2)	0.0341
NIGHT 57	71	9.9 (21.3)	71	9.5 (18.4)	-0.3 (3.0)	0.8624	75	7.2 (16.0)	-3.2 (3.0)	0.6482
NIGHT 85	70	12.2 (21.7)	69	10.8 (21.8)	-1.5 (3.3)	0.4520	74	5.4 (13.2)	-7.2 (3.2)	0.1077

There was no significant difference between Doxepin 3 mg and placebo on the objective number of awakenings after sleep onset (NAASO) at night 1 (3 mg vs. placebo p=0.2930) or night 85 (p=0.1135) (or nights 29 or 57). This, together with the (b) (4) results, may raise questions about the sponsor's claim of an effect towards (b) (4), which they based on the sleep efficiency endpoint results.

3.1.6 Study 509

Title: A Phase III, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Multicenter, Outpatient Study to Assess the Efficacy And Safety of Doxepin Hcl in Elderly Patients with Primary Sleep Maintenance Insomnia

The first subject was enrolled on 20 January 2006 and the last subject completed on 11 September 2006. All Investigators in this study were based in the United States.

3.1.6.1 Study Design and Statistical Methods

Objectives

Primary Objective: To evaluate the efficacy of doxepin HCl (doxepin) 6 mg relative to placebo when administered nightly for up to four weeks in elderly subjects with primary sleep maintenance insomnia.

Secondary Objective: To evaluate the safety of doxepin 6 mg relative to placebo when administered nightly for up to four weeks in elderly subjects with primary sleep maintenance insomnia.

Final Protocol Amendment (dated Feb 13, 2006)

Eligible patients will be randomly assigned to one of the following dose groups in a 1:1 ratio:

- doxepin HCl 6 mg; or
- placebo.

Primary Efficacy Endpoint

The primary efficacy variable is sTST at Week 1. The null hypothesis to be tested is that the sTST values for the two treatments are equal. The alternative hypothesis is that the sTST values are different for doxepin HCl 6 mg compared to placebo. Hypothesis tests for the comparison of doxepin HCl at 6 mg versus placebo will be analyzed using an analysis of covariance (ANCOVA) model that includes the main effects for treatment and center with the baseline sTST as a covariate. Baseline sTST is defined as the mean sTST from the nights that the patient self-administered the placebo during the Placebo Lead-In Period.

If the data are inconsistent with the assumptions of ANCOVA, an appropriate transformation of the data may be performed or appropriate non-parametric tests will be used to evaluate the differences between treatment groups.

Secondary Efficacy Endpoints

Continuous secondary efficacy variables (LSO, sNAASO, sWASO, ISI and the Clinical Global Impressions assessed by the clinician for severity of illness) will be analyzed using the same methods used to compare the sTST values using an ANCOVA model as appropriate.

LSO is expected to be log-normally distributed and will be transformed prior to analysis by taking the natural logarithm. The log-transformation will be performed after averaging values

reported via the IVRS from the nights that the patient self-administered study drug for the week. Any values of zero will be set to 0.25 minutes to permit calculation of the log-transformation.

Sensitivity analyses using methods to impute missing data will also be performed. Missing weekly averages will be imputed using both the last observation carried forward (LOCF) and the baseline observation carried forward (BOCF) methods. The last nonmissing weekly average will be carried forward for the LOCF method and the baseline weekly average will be carried forward for the BOCF method. These sensitivity analyses will be performed for sTST, LSO, and sWASO at Weeks 1, 2, 3, and 4 using the ITT analysis set.

Sample Size Determination

Approximately 240 patients will be randomized. The sample size for this study was based on results of Study SP-0402, a phase II cross-over study conducted in 76 elderly patients with primary sleep maintenance insomnia. Results from the ITT analysis set in this study showed that the mean difference in sTST between the doxepin HCl 6 mg and placebo groups was 31 minutes, with an estimated pooled standard deviation of 68 minutes. Using these estimates, 120 subjects per arm provides greater than 90% power to detect a difference in sTST between the doxepin HCl 6 mg and placebo groups of at least 30 minutes, using a two-sided two-sample t-test at the 5% level of significance.

3.1.6.2 Patient Disposition

A summary of disposition for all randomized subjects is provided in Table 38. A total of 525 subjects were screened for this study. Of the 255 randomized subjects, 254 subjects received double-blind study drug and were included in the Safety Analysis Set. Overall, 237 subjects (93%) completed the study. Of the 18 subjects (7%) who withdrew from the study, one subject discontinued after randomization but before taking a dose of double-blind study drug. Two subjects withdrew from the study due to an AE. One subject in the doxepin 6 mg group withdrew from the study due to hypoacusis of the left ear and tinnitus. The most frequently reported reason for discontinuation in both treatment groups was consent withdrawn (6% in the placebo group and 3% in the doxepin 6 mg group).

Table 38 Study 509: Patient Disposition, All Randomized Patients

Disposition	Placebo (N=125) n (%)	Doxepin 6 mg (N=130) n (%)	Total (N=255) n (%)
Completed the Study	113 (90%)	124 (95%)	237 (93%)
Withdrew from the Study	12 (10%)	6 (5%)	18 (7%)
Adverse Event	1 (1%)	1 (1%)	2 (1%)
Protocol Violation	1 (1%)	1 (1%)	2 (1%)
Noncompliance	2 (2%)	0 (0%)	2 (1%)
Consent Withdrawn	8 (6%)	4 (3%)	12 (5%)
Lost to Follow-up	0 (0%)	0 (0%)	0 (0%)
Death	0 (0%)	0 (0%)	0 (0%)
Other	0 (0%)	0 (0%)	0 (0%)

3.1.6.3 Baseline Demographics

Demographic and other baseline characteristics, summarized by treatment group in Table 39, were similar between treatment groups. The study population was female (65%) and male (35%). The mean age was 72.5 years. Subjects were White (87%), Black/African American (7%), Hispanic (3%), Asian (2%), and Other (1%).

Table 39 Study 509: Demographic and other baseline characteristics, Safety Set

Variable	Placebo (N=124)	Doxepin 6 mg (N=130)	Total (N=254)
Age (years)			
Mean (SD)	72.5 (5.85)	72.4 (5.95)	72.5 (5.89)
Median (Range)	71.0 (64–90)	71.5 (64–91)	71.0 (64–91)
Sex [n (%)]			
Male	48 (39%)	42 (32%)	90 (35%)
Female	76 (61%)	88 (68%)	164 (65%)
Race/Ethnicity [n (%)]			
White	108 (87%)	114 (88%)	222 (87%)
Black/African American	7 (6%)	10 (8%)	17 (7%)
Hispanic	5 (4%)	3 (2%)	8 (3%)
Asian	2 (2%)	3 (2%)	5 (2%)
Other	2 (2%)	0 (0%)	2 (1%)
Weight (kg)	n=123	n=130	n=253
Mean (SD)	77.1 (16.69)	77.4 (15.53)	77.2 (16.07)
Median (Range)	78.5 (36–133)	75.8 (46–135)	76.3 (36–135)
Height (cm)	n=124	n=130	n=254
Mean (SD)	166.3 (10.29)	165.4 (8.73)	165.8 (9.51)
Median (Range)	165.1 (142.2–201.0)	165.1 (147.3–190.5)	165.1 (142.2–201.0)
BMI (kg/m²)	n=123	n=130	n=253
Mean (SD)	27.7 (5.02)	28.2 (4.94)	28.0 (4.97)
Median (Range)	27.1 (18.0–50.3)	27.9 (18.7–43.5)	27.7 (18.0–50.3)

3.1.6.4 Sponsor’s Results

Primary Efficacy Variable – Subjective Total Sleep Time (sTST) at Week 1

A total of 32 of the 34 study centers randomized subjects into the study. Of these, data from 12 centers (06, 53, 61, 64, 66, 74, 75, 79, 80, 83, 85, and 87) with low enrollment (i.e., less than five subjects in the ITT Analysis Set) were pooled to form one pseudo-center, as described in the SAP.

The primary efficacy variable was sTST at Week 1.

A summary of the primary efficacy variable, sTST, at baseline and at Week 1 by treatment group using the ITT Analysis Set is provided in Table 40. At baseline, the mean sTST values were slightly higher in the placebo group (293.5 minutes) than in the doxepin 6 mg group (283.1 minutes). At Week 1, there was a statistically significant increase ($p < 0.0001$) in the mean sTST value for doxepin 6 mg compared with placebo. The LS mean sTST value was 28.6 minutes longer in the doxepin 6 mg group compared with the placebo group. An additional sensitivity analysis was performed on a subset of the ITT Analysis Set that excluded site 04, which incorrectly instructed subjects to round the IVRS data during the Placebo Lead-in Period to the nearest 15 minutes. Similar results were observed using the PP Analysis Set and sensitivity analyses.

Table 40 Study 509: Primary Efficacy Variable – Subjective Total Sleep Time – at Baseline and Week 1: ITT Analysis Set

sTST (minutes)	Placebo (N=124)	Doxepin 6 mg (N=130)
Baseline¹	n=124	n=130
Mean (SD)	293.5 (49.09)	283.1 (49.96)
Median (Range)	300.0 (107.1–385.7)	288.2 (158.6–377.1)
Week 1¹	n=122	n=129
Mean (SD)	316.7 (56.22)	335.2 (61.20)
Median (Range)	318.9 (156.0–458.6)	336.4 (201.4–475.7)
Difference of LS Mean (Std. Err.)		28.6 (5.39)
95% CI of LS Mean Difference		(18.0, 39.3)
p-value ³		p<0.0001
Week 2¹	n=118	n=127
Mean (SD)	328.2 (54.28)	332.3 (64.63)
Median (Range)	334.6 (192.0–492.0)	335.0 (175.7–497.1)
Difference of LS Mean (Std. Err.)		14.2 (5.61)
95% CI of LS Mean Difference		(3.1, 25.2)
p-value ³		p=0.0121
Week 3¹	n=114	n=125
Mean (SD)	331.7 (60.87)	341.4 (62.06)
Median (Range)	336.2 (181.4–475.7)	334.3 (216.4–479.3)
Difference of LS Mean (Std. Err.)		18.2 (6.09)
95% CI of LS Mean Difference		(6.2, 30.2)
p-value ³		p=0.0031
Week 4¹	n=109	n=122
Mean (SD)	336.4 (64.71)	346.1 (66.44)
Median (Range)	345.0 (153.8–468.3)	341.8 (201.4–496.3)
Difference of LS Mean (Std. Err.)		21.1 (6.63)
95% CI of LS Mean Difference		(8.0, 34.1)
p-value ³		p=0.0017
Average Across Weeks 1 Through 4²	n=108	n=122
Mean (SD)	330.5 (54.92)	338.9 (58.88)
Median (Range)	337.5 (182.1–468.8)	333.5 (226.7–463.7)
Difference of LS Mean (Std. Err.)		18.3 (5.34)
95% CI of LS Mean Difference		(7.8, 28.9)
p-value ³		p=0.0007

Notes: Values of sTST >12 hours were excluded from the analysis.

¹ Baseline was the mean of the IVRS values reported during the Placebo Lead-in Period. Weeks 1, 2, 3, and 4 were defined as the mean of the IVRS reported values between study visits during the treatment period.

² Average across Weeks 1 through 4 was calculated as the average of all non-missing IVRS entries if at least four observations were available at each week.

³ p-value comparing treatments was determined from an ANCOVA model that included main effects for treatment and center with the baseline value as a covariate.

Latency to Sleep Onset at Week 1

The key secondary efficacy variable was subjective Latency to Sleep Onset (LSO) during Week 1. A summary of LSO at baseline and Week 1 by treatment group using the ITT Analysis Set is provided in Table 41. At baseline, the LSO values were slightly lower in the placebo group. The geometric LS mean LSO at Week 1 in the doxepin 6 mg group was not statistically significantly different than placebo (p=0.1547). Similar results were observed using the PP Analysis Set and sensitivity analyses.

Numerical decreases in LSO from baseline were observed throughout the treatment period in the doxepin 6 mg group and the placebo group. There were no statistically significant differences

between the two treatment groups at any noted timepoint using the ITT Analysis Set. Similar results were observed using the PP Analysis Set, as well as with the sensitivity analyses.

Table 41 Study 509: Latency to Sleep Onset by Week: ITT Analysis Set (OC)

LSO (minutes)	Placebo (N=124)	Doxepin 6 mg (N=130)
Baseline¹	n=124	n=130
Mean (SD)	77.3 (37.81)	85.1 (45.53)
Geometric Mean	69.5	75.3
Median (Range)	68.5 (22.9–245.0)	72.7 (25.0–246.4)
Week 1²	n=122	n=129
Mean (SD)	68.0 (41.66)	67.0 (33.28)
Geometric Mean	59.1	58.9
Median (Range)	59.6 (17.0–287.1)	61.4 (14.4–159.3)
Geometric LS Mean ³	60.1	55.8
LS mean ratio		0.9
95% CI of LS mean ratio		(0.8, 1.0)
p-value ⁴		p=0.1547
Week 2²	n=118	n=127
Mean (SD)	59.0 (29.19)	69.7 (45.92)
Geometric Mean	52.2	58.1
Median (Range)	53.6 (12.1–161.4)	56.4 (10.7–295.7)
Geometric LS Mean ³	53.3	55.5
LS mean ratio		1.0
95% CI of LS mean ratio		(0.9, 1.2)
p-value ⁴		p=0.4884
Week 3²	n=114	n=125
Mean (SD)	60.7 (34.53)	65.3 (43.68)
Geometric Mean	52.1	53.4
Median (Range)	53.6 (15.8–190.0)	54.0 (11.3–268.0)
Geometric LS Mean ³	52.3	49.8
LS mean ratio		1.0
95% CI of LS mean ratio		(0.8, 1.1)
p-value ⁴		p=0.4635
LSO (minutes)	Placebo (N=124)	Doxepin 6 mg (N=130)
Week 4²	n=109	n=122
Mean (SD)	57.5 (35.00)	63.2 (42.40)
Geometric Mean	49.0	51.4
Median (Range)	51.4 (11.6–216.0)	55.7 (7.5–300.0)
Geometric LS Mean ³	50.9	49.4
LS mean ratio		1.0
95% CI of LS mean ratio		(0.8, 1.1)
p-value ⁴		p=0.6629

3.1.6.5 Reviewer’s Results

This reviewer verified the sponsor’s analyses of subjective TST and WASO. Table 42 shows the results for Observed Cases (the primary analysis population) as well as analyses using LOCF imputation of missing data for patients that dropped out. The Doxepin 6 mg group was significantly reduced in terms of the subjective WASO over each week. The consistency of the p-values for the OC and LOCF analyses suggests that it is not likely that the missing data has biased the OC results.

Table 42 Study 509: Subjective WASO analyses for OC and ITT-LOCF populations by Week

POP	TIME	PLACEBO		6 MG		6 MG VS. PLACEBO	
		N	Mean (S.D.)	N	Mean (S.D.)	Difference LS Mean (S.E.)	p-value
OC	Baseline	125	111.3 (47.1)	130	116.5 (49.1)	5.9 (6.0)	0.3313
	Week 1	122	97.4 (50.2)	129	79.1 (49.0)	-22.2 (4.8)	<0.0001
	Week 2	118	85.1 (50.4)	127	75.9 (47.0)	-12.6 (5.1)	0.0145
	Week 3	114	82.4 (49.7)	125	70.4 (46.8)	-15.8 (4.9)	0.0016
	Week 4	108	78.9 (56.5)	122	66.5 (43.9)	-16.8 (5.5)	0.0026
LOCF	Week 2	122	86.2 (50.3)	129	75.9 (47.0)	-13.8 (5.1)	0.0071
	Week 3	122	85.0 (50.7)	129	71.2 (46.9)	-17.8 (4.9)	0.0004
	Week 4	122	83.7 (56.2)	129	67.8 (43.9)	-19.1 (5.3)	0.0003

This reviewer also verified the sponsor’s analyses of subjective Latency to Sleep Onset. Table 43 shows the results for Observed Cases (the primary analysis population) as well as analyses using

LOCF imputation of missing data for patients that dropped out. Differences in LSO were not statistically significant at any week. The consistency of the p-values for the OC and LOCF analyses suggests that it is not likely that the missing data has biased the OC results.

Table 43 Study 509: Subjective LSO analyses for OC and ITT-LOCF populations by Week

POP	TIME	PLACEBO			6 MG			
		N	Mean (S.D.)	Geometric Mean	N	Mean (S.D.)	Geometric Mean	6 MG vs. placebo p-value
OC	Baseline	125	4.24 (0.46)	69.3	130	4.32 (0.49)	75.3	0.1244
	Week 1	122	4.08 (0.51)	59.1	129	4.08 (0.52)	58.9	0.1547
	Week 2	118	3.96 (0.51)	52.2	127	4.06 (0.61)	58.1	0.4884
	Week 3	114	3.95 (0.56)	52.1	125	3.98 (0.64)	53.4	0.4635
	Week 4	109	3.89 (0.58)	49.0	122	3.94 (0.67)	51.4	0.6629
LOCF	Week 2	122	3.96 (0.53)	52.6	129	4.07 (0.61)	58.6	0.4790
	Week 3	122	3.97 (0.58)	53.2	129	3.99 (0.65)	54.0	0.3596
	Week 4	122	3.93 (0.60)	50.8	129	3.96 (0.68)	52.6	0.5440

Other Endpoints

There was no significant difference between Doxepin 6 mg and placebo on the subjective number of awakenings after sleep onset (NAASO) for any week, e.g., week 1 (p=0.1025), week 4 (p=0.9175). This may raise questions about the sponsor's claim of an effect towards (b) (4)

3.2 Evaluation of Safety

Please see the medical review for the Evaluation of Safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Note that in the following tables p-values are provided to give an idea of the size of the observed differences; however, the p-values should be viewed as exploratory and they have not been adjusted for multiplicity.

4.1.1 Gender

In study 501, 73% of randomized patients were female and 27% were male.

A statistical test for a differential effect of treatment depending on gender was not significant $p=0.285$ and both groups' differences from placebo on WASO reached the nominal significance level.

Table 44 Study 501: Gender Subgroup Analyses of Objective WASO on Night 1

TREAT GROUP	MALE				FEMALE				ALL	
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	Mean
P	22	80.77 (44.37)	83.41 (63.28)	.	50	80.77 (44.37)	59.33 (42.02)	.	72	66.7 (50.5)
3	17	69.78 (35.11)	37.59 (31.43)	<0.001	58	67.16 (33.38)	42.55 (31.72)	0.001	75	41.4 (32.3)
6	21	78.31 (38.16)	45.86 (27.08)	0.001	52	59.60 (29.75)	32.41 (24.98)	<0.001	73	36.3 (26.1)

Interaction test $p=0.285$

In study 503, 70% of randomized patients were female and 30% were male.

A statistical test for a differential effect of treatment depending on gender was not significant $p=0.7628$ and both groups' differences from placebo on WASO reached the nominal significance level except for males on 1mg, but this may be due to the lower power.

Table 45 Study 503: Gender Subgroup Analyses of Objective WASO on Night 1

TREAT GROUP	MALE				FEMALE				ALL	
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	Mean
P	33	129.9 (44.4)	122.5 (46.2)	.	48	129.9 (44.4)	99.5 (43.9)	.	81	108.9 (45.7)
1	27	129.6 (36.4)	109.8 (51.2)	0.168	50	114.9 (33.4)	82.0 (42.1)	0.021	77	91.7 (46.4)
3	25	124.3 (23.9)	92.4 (41.2)	0.017	57	115.1 (29.6)	66.6 (33.8)	<0.001	82	74.5 (36.9)

Interaction test $p=0.7628$

In study 509, 65% of randomized patients were female and 35% were male.

A statistical test for a differential effect of treatment depending on gender was not significant $p=0.7689$ and in both gender groups the 6 mg group difference from placebo on WASO reached the nominal significance level.

Table 46 Study 509: Gender Subgroup Analyses of Subjective WASO over Week 1

TREAT GROUP	MALE				FEMALE				ALL	
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	Mean
P	47	113.6 (44.7)	99.6 (54.6)	.	75	113.6 (44.7)	96.0 (47.5)	.	122	97.4 (51.0)
6	42	125.3 (52.3)	85.2 (55.0)	0.004	87	112.3 (47.2)	76.1 (45.9)	0.001	129	79.1 (49.6)

Interaction test $p=0.7689$

Over all randomized patients in the transient insomnia study (502), 55% were female and 45% were male. A statistical test for a differential effect of treatment depending on gender was not significant $p=0.8458$ and in both gender groups the 6 mg group difference from placebo in terms of LPS reached the nominal significance level.

Table 47 Study 502: Gender Subgroup Analyses of Objective LPS on Night 1

TREAT Group	Male				FEMALE				ALL	
	N	MEAN Log LPS (SD)	Geom Mean	Pvalue	N	MEAN Log LPS (SD)	Geom Mean	Pvalue	N	MEAN Log LPS (SD)
P	134	3.0 (1.0)	20.2	.	148	3.1 (1.0)	21.1	.	282	3.1 (1.0)
6	122	2.7 (1.0)	14.5	0.004	160	2.7 (0.9)	14.2	0.001	282	2.7 (0.9)

Interaction test $p=0.8458$

4.1.2 Race

In study 501, randomized patients' ethnicities were White (48%), Black (33%), Hispanic (16%), and Others (3%). A statistical test for a differential effect of treatment in terms of the objective WASO, depending on race, was not significant $p=0.2968$.

Table 48 Study 501: Race Subgroup Analyses of Objective WASO on Night 1

TREAT Group	Hisp				Af Amer				Cauc				Other			
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value
P	11	48.4 (22.8)	38.9 (15.9)	.	24	48.4 (22.8)	71.2 (61.6)	.	35	48.4 (22.8)	70.5 (46.1)	.	2	48.4 (22.8)	99.5 (75.7)	.
3	15	63.3 (37.1)	42.5 (33.9)	0.627	26	69.6 (33.9)	39.9 (27.8)	<0.001	33	68.2 (33.0)	42.4 (34.4)	0.001	1	71.5 (.)	31.0 (.)	0.186
6	10	62.9 (26.8)	46.4 (39.3)	0.921	21	68.2 (42.3)	30.8 (22.3)	<0.001	39	65.4 (30.5)	36.0 (24.1)	<0.001	3	44.1 (15.3)	45.2 (26.4)	0.356

Interaction test $p=0.2968$

In study 503, randomized patients' ethnicities were White (80%), Black (9%), Hispanic (9%), and Others (2%). The sample sizes were too small in the non-white ethnicities to allow for any reliable comparisons of efficacy between different ethnicities. The suggestion that the treatment effect differs by race is probably due to a large placebo response in the Hispanic and Other groups. These can likely be attributed to the extra variability of means associated with small sample sizes rather than to a differential effect of treatment according to ethnicity.

Table 49 Study 503: Race Subgroup Analyses of Objective WASO on Night 1

TREAT Group	Hispanic				Black				Cauc.				Other			
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value
P	4	110.3 (29.0)	52.0 (11.1)	.	6	110.3 (29.0)	110.2 (42.8)	.	67	110.3 (29.0)	113.9 (45.6)	.	4	110.3 (29.0)	78.5 (39.0)	.
1	8	120.3 (37.5)	105.3 (50.0)	0.058	5	146.8 (32.5)	126.1 (68.6)	0.949	63	117.8 (34.7)	86.1 (43.3)	<0.001	1	130.3 (.)	171.5 (.)	0.066
3	9	115.1 (16.9)	68.8 (38.2)	0.626	10	121.1 (30.9)	64.6 (33.0)	0.036	63	117.8 (29.3)	76.8 (38.8)	<0.001	0			.

Interaction test $p=0.0111$

Overall in study 509 ethnicities were White (87%), black (7%), Hispanic (3%), Asian (2%), and Other (1%). A statistical test for a differential effect of treatment in terms of the subjective WASO, depending on race, was not significant $p=0.2078$.

Table 50 Study 509: Race Subgroup Analyses of Subjective WASO over Week 1

TREAT	GROUP Cauc.				GROUP Other				ALL	
	N	BL MEAN (SD)	Mean	p-value	N	BL MEAN (SD)	Mean	p-value	N	Mean
P	106	113.5 (46.2)	98.7 (50.9)	.	16	113.5 (46.2)	88.4 (46.0)	.	122	97.3 (51.0)
6	113	116.0 (50.4)	76.7 (48.9)	<0.001	16	119.8 (40.0)	95.9 (48.0)	0.685	129	79.1 (49.4)

Interaction test $p=0.2078$

In the transient insomnia study (502) ethnicities over all randomized patients were White (50%), Black (15%), Hispanic (32%) and Other (3%). A statistical test for a differential effect of treatment in terms of the objective LPS, depending on race, was not significant $p=0.7674$.

Table 51 Study 502: Race Subgroup Analyses of Objective LPS on Night 1

TREAT Group	Hisp.			Afr. Ameri			Cauc.			Other			ALL	
	N	MEAN Log LPS (SD) Geom Mean	Pvalue	N	MEAN Log LPS (SD) Geom Mean	Pvalue	N	MEAN Log LPS (SD) Geom Mean	Pvalue	N	MEAN Log LPS (SD) Geom Mean	Pvalue	N	MEAN Log LPS (SD)
P	87	3.1 (1.0) 22.0	.	39	2.8 (1.0) 16.7	.	151	3.1 (1.0) 21.6	.	5	2.3 (1.5) 10.3	.	282	3.0 (1.0)
6	93	2.6 (0.9) 13.8	0.001	45	2.5 (1.0) 12.2	0.124	132	2.8 (0.9) 16.3	0.01	12	2.2 (0.6) 8.7	0.869	282	2.7 (0.9)

Interaction test $p=0.7674$

4.1.3 Age

Overall, the observed treatment group differences did not appear to depend significantly on patient age.

In study 501, overall, the mean age was 44.5 and ages ranged from 18 to 65. A statistical test for a differential effect of treatment in terms of the objective WASO, depending on age, was not significant $p=0.8012$.

Over all randomized patients in the elderly study 503 the average age was 71.4 and ages ranged from 64 to 93. A statistical test for a differential effect of treatment in terms of the objective WASO, depending on age, was not significant $p=0.4091$.

In the elderly study 509, which utilized only subjective measures, overall the average age was 72.5 and ages ranged from 64 to 91.

A statistical test for a differential effect of treatment in terms of the subjective WASO, depending on age, was not significant $p=0.6653$.

In the transient insomnia study (502) over all randomized patients the average age was 35.5 and ages ranged from 25 to 55. A statistical test for a differential effect of treatment in terms of the subjective LPS, depending on age, was not significant $p=0.5360$.

4.2 Other Special/Subgroup Populations

4.2.1 Treatment Group Differences by Site

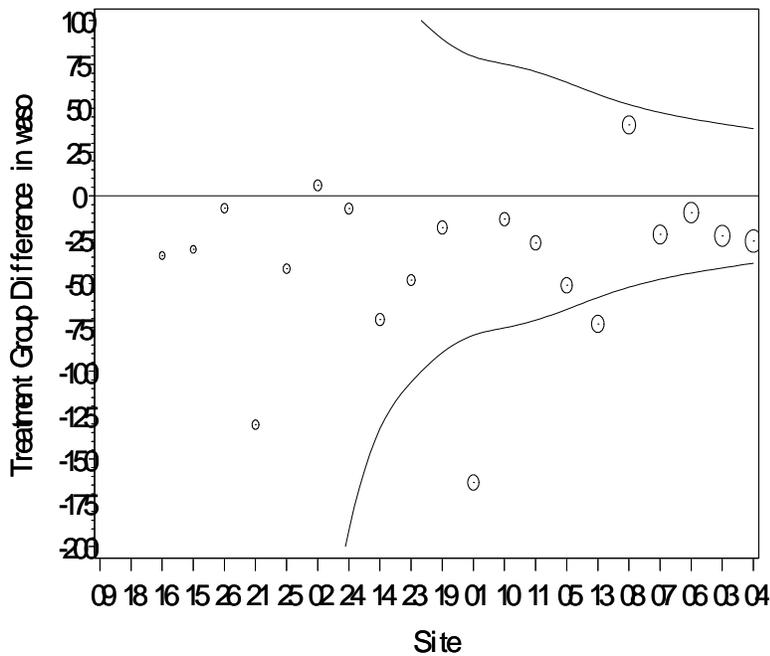
The following graphs show treatment group mean differences by Site. The size of the plotting symbol in the graph indicates the relative sample size at a given site compared to the other sites and the sites are ordered along the horizontal axis from smallest to largest.

The curves on the graph permit roughly judging the nominal significance of the observed center specific effect (adjusted for sample size) in an exploratory fashion¹.

In study 501, 22 sites randomized patients. Site total sample size ranged from 1 to 34; the average size was 13. For 18 out of the 20 sites that had at least 1 patient per group (6 mg and placebo) the mean difference in objective WASO between 6 mg and placebo numerically favored Doxepin 6 mg. The p-value for a test for differential treatment effects by pooled site was 0.083. This p-value may be driven by the treatment group difference in site 1, which appears to be an outlier, because the test for interaction p-value increases to 0.378 if site 1 is excluded. Aside from site 1 the treatment group differences by site were relatively consistent. Among the bigger sites, siteid's 13 (Ereshefsky N=15) and 4 (Lankford, N=25) in that order had the biggest treatment group differences on objective WASO. Excluding site 1 data did not alter the significance of the treatment group differences in objective WASO on night 1.

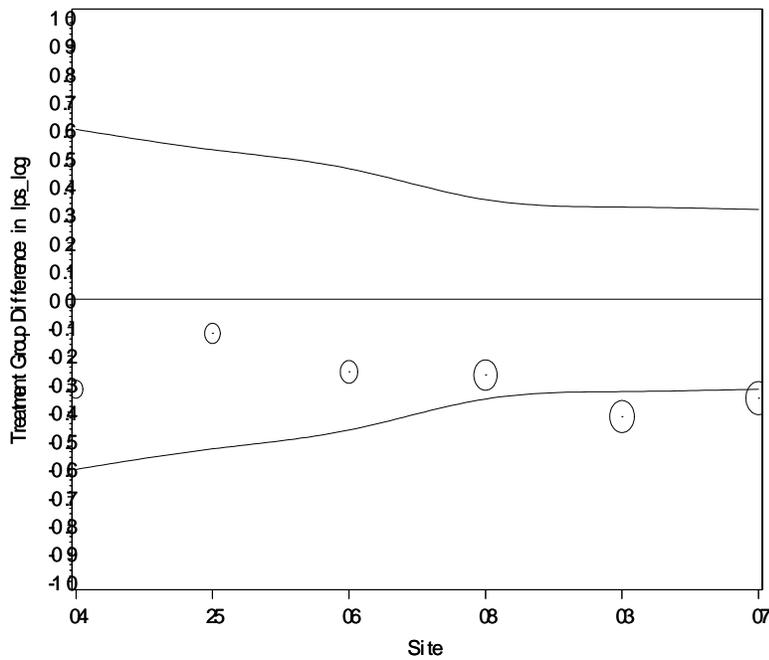
¹ The curves were determined by taking twice the square root of the estimate of the residual error variance (from the main analysis) and adjusting that according to the group's sample size within each site.

Figure 3 Study 501: 6 mg and Placebo Estimated Night 1 Difference in Objective WASO by Site



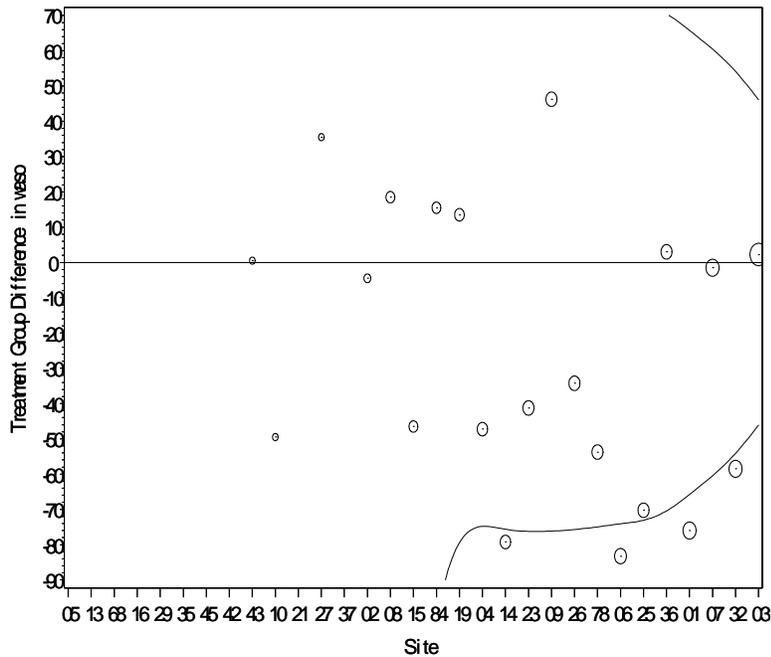
In study 502 (transient insomnia) 6 sites randomized patients. The average total sample size was 141 and sizes ranged from 63 to 216. All 6 sites numerically favored Doxepin 6 mg over placebo on Night 1 in terms of LPS (Figure 4). The p-value for a test for differential treatment effects by pooled site was 0.891. Scharf (N=68), the investigator for site 6, had the biggest effect on WASO but was fourth biggest out of six on the latency to persistent sleep (LSP) endpoint. Site 3, Dr. Hull (N=135) was largest on LPS and second largest on WASO. However, excluding site 3 did not alter the significance of the treatment difference.

Figure 4 Study 502: 6 mg and Placebo Estimated Night 1 Difference in Objective Log (LPS) by Site



In study 503, in the elderly, 31 sites randomized patients. Site total sample size ranged from 1 to 37; the average size was 10. For 13 out of the 21 sites that had at least 1 patient per group (3 mg and placebo) the mean difference in objective WASO between 3 mg and placebo numerically favored Doxepin 3 mg. The p-value for a test for differential treatment effects by pooled site was 0.543. Site 6 (Scharf) had the biggest effect on WASO. Site 1(Corser N=15), 25 (Orr N=13), and 32(Gottfried, N=16) also had big effects in that order. Excluding these sites individually did not alter the significance of the treatment group differences in objective WASO on Night 1.

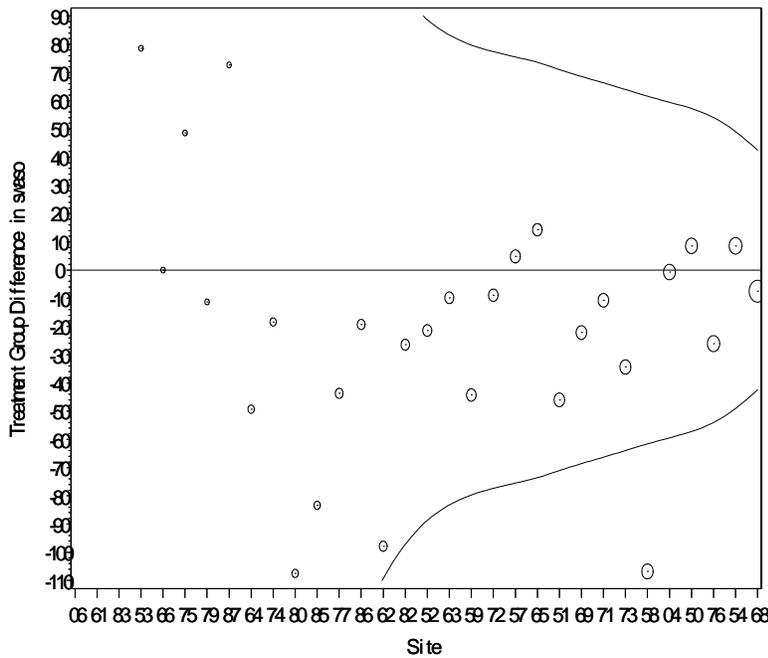
Figure 5 Study 503: 3 mg and Placebo Estimated Night 1 Difference in Objective WASO by Site



In the elderly study which had subjective measures only (study 509), 32 sites randomized patients. Site total sample size ranged from 0 to 42; the average size was 12. For 21 out of the 29 sites that had at least 1 patient per group the mean difference in subjective WASO over week 1 numerically favored Doxepin. Sites 76 (Segal, N=15), 50 (Anderson, N=14), 58 (Essink, N=13), 71 (Merideth, N=12) had the biggest treatment differences on the sponsor's designated primary, subjective total sleep time, ordered from smallest to largest in terms of sample size and size of the difference. The p-value for a test for differential treatment effects by pooled site on sWASO was 0.946.

Site 58 (Essink, N=13) had what appeared to be a big outlying treatment group difference on subjective WASO. Site 76 also had a moderately big treatment group difference. However, excluding these sites individually did not alter the significance of the treatment group differences in subjective WASO over Week 1.

Figure 6 Study 509: 3 mg and Placebo Estimated Week 1 Difference in Subjective WASO by Site



5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The sponsor’s prespecified analysis plans for dealing with type I error issues associated with multiple endpoints and multiple timepoints often differed from the approach preferred by the FDA Neurology division. The division recommended testing the objective WASO, followed by the subjective WASO, the objective LPS, and finally the subjective LSO. In addition for each efficacy measure they recommended starting at the latest time and working backwards to the first time until an insignificant result was observed. When there were multiple doses in a study this procedure would start with the high dose first. The results of this procedure can be determined from Table 52. In study 501 the 6 mg group was significantly improved compared to placebo in terms of the objective WASO at each visit (1st of two nights). However, the 6 mg group was not statistically significantly better than the placebo group in terms of the subjective WASO at night 29 (1st night of the last visit). Therefore, in order to control the experiment wise type I error at 0.05 level testing must stop with this test, i.e., the results on the LPS and sLSO endpoints, as well as the results for 1mg vs. placebo comparisons, can only be considered exploratory. This reviewer found that if the sponsor had chosen to perform the comparison on the 2nd night of each visit or the average of nights 1 and 2 instead of on the first night only then the 6 mg group would

have won at each time in terms of the subjective WASO. However, the choice of which night's data to base the hypothesis on introduces another layer of multiplicity if we consider something different than the sponsor prespecified. Even if they had chosen differently we see that they would not have won on the LPS because the 6mg vs. placebo difference was not significant at night 29. There is no evidence from any of the studies that there is a statistically significant effect on latency to persistent sleep beyond night 1.

Table 52 Summary of Key Analyses by Study

Study	Endpoint	Dose Group	P-Values as Compared to Placebo				
			Night 85	Night 57	Night 29	Night 15	Night 1
401(Phase 2 Crossover)	WASO	6					<0.0001
	LPS	6					0.0397
402 (Phase 2 Crossover/ Elderly)	WASO	6					<0.0001
	LPS	6					0.1063
	WASO	6			0.0007	0.0011	<0.0001
		3			0.0173	0.0025	<0.0001
	sWASO	6			0.6282	0.2016	0.0004
		3			0.6483	0.1512	0.0003
	LPS	6			0.8643	0.5921	0.0009
		3			0.7995*	0.2271*	0.0058
	sLSO	6			0.6511*	0.1451	0.0492
		3			0.2365*	0.9071*	0.1259
502 (Transient Insomnia)	WASO	6					<.0001
	sWASO	6					0.0063
	LPS	6					<0.0001
	LSO	6					<0.0001
503 (Elderly)	WASO	3	<.0001	0.0029	0.0005	0.0069	<0.0001
		1	0.0330	0.1662	0.0878	0.1945	0.0053
	sWASO	3	0.0153	0.5627	0.0296	0.0729	0.0561
		1	0.0037	0.7417	0.0531	0.8571	0.8497
	LPS	3	0.0286*	0.0522*	0.5422*	0.8388	0.1079
		1	0.6493	0.1870*	0.1268*	0.8046	0.5733*
	sLSO	3	0.8479	0.9931	0.6544*	0.916	0.0860
		1	0.2826	0.9631*	0.1798*	0.3567*	0.2304*
509 (Elderly/ Subjective Only)	sWASO	6		0.0026 (Week 4)	0.0016 (Week 3)	0.0145 (Week 2)	<0.0001 (Week 1)
	sLSO	6		0.6629 (Week 4)	0.4635 (Week 3)	0.4884* (Week 2)	0.1547 (Week 1)

Note: The empty cells reflect the different lengths of study. In studies 401 and 402 the night 1 results are actually the results for the average of nights 1 and 2 as pre-specified by the sponsor in the analysis plan. For study 509 which analyzed weeks instead of individual nights the time corresponding to the analysis is displayed in the cell.

* numerically favors placebo

For study 501, the sponsor did not consider nights after the first night to be key hypotheses, i.e., to include them in the set of hypotheses which would have the chance of a single type I error over all hypotheses in it protected at 0.05. The sponsor named key secondary endpoints but did not state unambiguously that they were to be tested in order until an insignificant result was obtained.

For study 503 the sponsor specified a clear hierarchy for testing. The subjective Total Sleep Time over night 1 was the next endpoint after the objective WASO was tested at each night. Because the comparison between 3 mg and placebo on the sTST over Night 1 was not significant ($p=0.0865$) no claims can be made on the lower endpoints in the hierarchy (LPS and Sleep Efficiency). Note that the 3mg vs. placebo comparison of LPS on night 1 was not significant ($p=0.1079$) either which is even more reason that no claims of an effect on Sleep Efficiency are possible.

In summary, study 501 in non-elderly adults provides some evidence for the superior efficacy of the 6 mg dose compared to placebo for sleep maintenance as measured by objective WASO. The differences between 6 mg and placebo in terms of subjective WASO were not consistently significant. They were not significant at the sponsor's prespecified key timepoints, first night of each visit, but they were nominally significant at the second night of each visit as well as for the average of the two nights at each visit. Although, the 3 mg dose was also nominally significantly better compared to placebo for sleep maintenance as measured by objective WASO the multiplicity adjustment requires us to consider these results as exploratory because of the insignificant LPS results for the 6 mg dose that were before all 3 mg comparisons in the testing hierarchy. However, study 503 in elderly patients provides some evidence of efficacy of the 3 mg dose compared to placebo for sleep maintenance as measured by objective WASO. The 1 mg dose in study 503 has the same problem as the 3 mg dose in study 501, i.e., it is below some insignificant 3 mg comparisons in the testing hierarchy. In addition, the differences between 1 mg and placebo in terms of objective WASO at intermediate times between night 1 and night 85 were not nominally significant. Study 509 in the elderly only included subjective measures. The sponsor actually specified the Total Sleep Time as the primary measure in study 509 but if we believe the WASO to be a better measure we can also examine it because the Total Sleep Time results were statistically significant in favor of the Doxepin 6 mg group. The subjective WASO results were also positive. Therefore, study 509 provides some evidence of the superior efficacy of 6 mg over placebo for sleep maintenance as measured by the subjective Total Sleep Time or subjective WASO. Based on these considerations, overall, it seems that there may be sufficient evidence to support the efficacy of the 3 and 6 mg doses for sleep maintenance.

5.2 Conclusions and Recommendations

The clinical efficacy studies in this application seem to support an effect of the drug on sleep maintenance as measured by the Wake time after sleep onset. However, there were no consistent effects of the drug on latency to persistent sleep; no statistically significant differences between Doxepin and placebo in terms of LPS were observed beyond night 1 in any study. The sponsor's claim of an effect on (b) (4) based on the Sleep Efficiency endpoint is also questioned for two reasons. First, because after adjusting for multiple testing it is not possible to make a claim on Sleep Efficiency without inflating the experimentwise type I error because Sleep Efficiency was lower in the testing order than LPS (e.g., in study 503). Second, there was not a statistically significant effect of the drug compared to placebo on the Wake Time After Sleep endpoint or the Number of Awakenings after Sleep Onset endpoint.

**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
12/10/2008 01:08:42 PM
BIOMETRICS

Kun Jin
12/11/2008 11:30:43 AM
BIOMETRICS

Kooros Mahjoob
12/16/2008 03:02:52 PM
BIOMETRICS