CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

205677Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA Serial Number:	NDA 205-677 SN000
Drug Name:	Tasimelteon
Indication(s):	Non-24 Hour Disorder in Totally Blind Individuals
Applicant:	Vanda Pharmaceuticals Inc.
Date of Submission	5/31/2013
Review Priority:	Priority
Biometrics Division:	Division of Biometrics I
Statistical Reviewer:	Jingyu (Julia) Luan, Ph.D. (HFD-710)
Concurring Reviewers:	Kun Jin, Ph.D. (HFD -710) H.M. James Hung, Ph.D. (HFD -710)
Medical Division:	Division of Neurology Product (HFD -120)
Clinical Team:	Devanand Jillapalli, M.D. (HFD-120) Ronald Farkas, M.D. (HFD-120)
Project Manager:	Ms. Cathleen Michaloski (HFD-120)
Keywords: ANCOVA, Permutation ANCOVA	

Table of Contents

LIST OF TABLES **4**

LIST OF FIGURES **4**

1 EXECUTIVE SUMMARY **6**

1.1 SPONSOR'S EFFICACY ANALYSES 6
1.2 SUMMARY OF IMPORTANT EVENTS RELATED TO STATISTICAL ANALYSIS IN STUDY 3201 7
1.3 SUMMARY OF STATISTICAL ISSUES 7
1.4 REVIEWER'S EFFICACY ANALYSES 9

2 INTRODUCTION **10**

2.1 OVERVIEW 10
2.2 DATA SOURCES 11

3 STATISTICAL EVALUATION **11**

3.1 DATA AND ANALYSIS QUALITY 11
3.2 EVALUATION OF EFFICACY 11
3.2.1 Protocol VP-VEC-162-3201 (SET Study) 11
3.2.1.1 Study Objectives 11
3.2.1.2 Study Design 13
3.2.1.3 Efficacy Measures 15
3.2.1.4 Statistical Analysis Plan 16
3.2.1.5 Patient Disposition, Demographic and Baseline Characteristics 17
Disposition of Patients 17
Datasets Analyzed 18
Demographics and Baseline Characteristics 19
3.2.1.6 Sponsor's Primary Efficacy Results 20
3.2.1.7 Sponsor's Secondary Efficacy Results 22
3.2.2 Protocol VP-VEC-162-3203 (RESET Study) 23
3.2.2.1 Study Objectives 23
3.2.2.2 Study Design 23
3.2.2.3 Efficacy Measures 24
3.2.2.4 Statistical Analysis Plan 24
3.2.2.5 Patient Disposition, Demographic and Baseline Characteristics 25
Patients Disposition 25
Datasets Analyzed 26
Demographic and Other Baseline Characteristics 27
3.2.2.6 Sponsor's Primary Efficacy Results 30
3.2.2.7 Sponsor's Secondary Efficacy Results 31
3.2.3 Reviewer's Analysis 32
3.2.3.1 Summary of Important Events Related to Statistical Analysis in Study 3201 32
3.2.3.2 Summary of Statistical Issues 33
Study 3201 33
Study 3203 34
3.2.3.3 Graphic Presentation of Clinical Endpoints 34
3.2.3.4 The Twelve Patients Excluded from Efficacy Analysis by Sponsor for Study 3201 45
3.2.3.5 Results of ANCOVA Analysis for Clinical Endpoints on ITT Population 47
3.2.3.6 Permutation ANCOVA Test for Clinical Endpoints for ITT Population 49
3.3 EVALUATION OF SAFETY 51

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS **51**

4.1 GENDER, RACE, AGE AND GEOGRAPHIC REGION 51
4.1.1 Study 3201 (SET Study) 51
4.1.2 Study 3203 (RESET Study) 55

4.2 OTHER SPECIAL/SUBGROUP POPULATIONS 56

5 SUMMARY AND CONCLUSIONS 56

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 56
5.1.1 Sponsor's Efficacy Analyses 56
5.1.2 Summary of Important Events Related to Statistical Analysis in Study 3201 57
5.1.3 Summary of Statistical Issues 57
5.1.4 Reviewer's Efficacy Analyses 59
5.2 CONCLUSIONS AND RECOMMENDATIONS 60

List of Tables

Table 1: Non-24 Scale of Clinical Response (N24CRS) 15
Table 2: Summary of Analysis Datasets 18
Table 3: Demographics and Baseline Characteristics – All Randomized Patients 19
Table 4: Summary of Primary Efficacy Endpoints (ITT and Analysis Population) 21
Table 5: Summary of Secondary Efficacy Endpoints (ITT and Analysis Population) 22
Table 6: Analysis Population Summary -- All Enrolled Patients 27
Table 7: Demographic and Baseline Characteristics – All Enrolled Patients and ITT population 28
Table 8: Summary of Primary Efficacy Endpoint (ITT Population) 31
Table 9: Summary of Secondary Efficacy Endpoints (ITT Population) 32
Table 10: Summary of Important Events Related to Statistical Analysis in Study 3201 33
Table 11: The 12 Patients Excluded from Efficacy Analysis (Study 3201) 46
Table 12: Descriptive Statistics for the 12 Patients Excluded from Sponsor Analysis Population 47
Table 13: LS Mean in Hours and P-value from ANCOVA with Site (Study 3201, ITT, n=84) .. 48
Table 14: LS Mean in Hours and P-value from ANCOVA without Site (Study 3201, ITT, n=84) 48
Table 15: LS Mean in Hours and P-value from ANCOVA without Site (Study 3203, ITT, n=20) 48
Table 16: Descriptive Statistics for Clinical Endpoints (Study 3201, ITT, n=84) 49
Table 17: Descriptive Statistics for Clinical Endpoints (Study 3203, ITT, n=20) 50
Table 18: Permutation ANCOVA with Site (Study 3201, ITT, n=84) 50
Table 19: Permutation ANCOVA without Site (Study 3201, ITT, n=84) 51
Table 20: Results of Permutation ANCOVA without Site (Study 3203, ITT, n=20) 51
Table 21: Subgroup Analysis by Sex (Study 3201, ITT, n=84) 52
Table 22: Subgroup Analysis by Race (Study 3201, ITT, n=84) 53
Table 23: Subgroup Analysis for Age Group (Study 3201, ITT, n=84) 54
Table 24: Subgroup Analysis for Country (Study 3201, ITT, n=84) 54

List of Figures

Figure 1: Study Design (US) 14
Figure 2: Study Design (Germany) 14
Figure 3: Flow Diagram of Patients Disposition (All Patients) 17
Figure 4: Study Design 24
Figure 5: Flow Diagram of Patient Disposition (All Patients) 26
Figure 6: Histograms and CDF for LQ-nTST (Study 3201, FDA ITT, n=84) 35
Figure 7: Histograms and CDF for UQ-dTSD (Study 3201, FDA ITT, n=84) 36
Figure 8: Histograms and CDF for MoST (Study 3201, FDA ITT, n=84) 37
Figure 9: Histograms and CDF for CGIC (Study 3201, FDA ITT, n=84) 38
Figure 10: Histograms and CDF for nTST (Study 3201, FDA ITT, n=84) 38
Figure 11: Histograms and CDF for dTSD (Study 3201, FDA ITT, n=84) 40

Figure 12: Histograms and CDF for LQ-nTST (Study 3203, n=20) 41
Figure 13: Histograms and CDF for UQ-dTSD (Study 3203, n=20) 42
Figure 14: Histograms and CDF for MoST (Study 3203, n=20) 43
Figure 15: Histograms and CDF for nTST (Study 3203, n=20) 44
Figure 16: Histograms and CDF for dTSD (Study 3203, n=20) 45

1 EXECUTIVE SUMMARY

This original NDA submission includes two pivotal efficacy studies, Study 3201 (SET Study) and Study 3203 (RESET Study). For both studies, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the clinical endpoints. According to this reviewer's post-hoc exploratory analyses without multiplicity adjustment, five out of the six clinical endpoints yield nominal p-values less than 0.05 or approximately 0.05; thus, the two studies appear to suggest that tasimelteon 20 mg may be beneficial for Non-24 Hour Disorder in Totally Blind Individuals based on all the clinical endpoints except for night Total Sleep Time (nTST, the original primary endpoint). However, the results of these two studies should be interpreted with caution, since they are based on post-hoc exploratory analysis without multiplicity adjustment and thus it is unknown whether the overall type I error is properly controlled.

Study 3201 was a multicenter, randomized, double-masked, placebo-controlled, parallel study designed to evaluate the efficacy and safety of 20 mg of tasimelteon versus placebo in patients suffering from Non-24. Eighty-four patients (tasimelteon 42; placebo 42) were randomized to receive tasimelteon (20 mg/day) or placebo. Of these 84 patients who were randomized to study drug, 62 (73.8%) patients completed the Randomization Phase and 22 (26.2%) patients discontinued early. This study was conducted at investigative sites in the US and Germany. The study began with a Pre-Randomization Phase (~5-6 weeks) and was followed by either a Randomization Phase (~26 weeks) or an Open-Label Extension Phase (~26 weeks).

Study 3203 was a multicenter, randomized withdrawal, double-masked, placebo-controlled, parallel group study designed to evaluate the long-term maintenance effect and safety of 20 mg of tasimelteon versus placebo in patients with Non-24. Every patient who enrolled in Study 3203 had previously been screened in 3201. Patients who met the inclusion criteria and who had previously participated in, or were screened for Study 3201, were eligible to participate. The study had 2 phases: a Pre-Randomization Phase (consisting of an Open-label tasimelteon Run-in Phase [~6 weeks] and a τ Estimation Phase [~ 6 weeks]), and a Randomized Withdrawal Phase (~8 weeks). Twenty patients in the US were randomized into the study and they all completed the Randomized Withdrawal Phase.

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the clinical endpoints, instead of a melatonin-based biomarker proposed by the sponsor. The sections below present sponsor's efficacy analyses, summary of important events related to statistical analysis in Study 3201, summary of statistical issues, and this reviewer's efficacy analyses.

1.1 Sponsor's Efficacy Analyses

Study 3201:

The primary efficacy endpoints were the following:

- The entrainment of the circadian melatonin rhythm as measured by urinary aMT6s (Entrainment is a melatonin-based biomarker and is defined as having a post-baseline τ value less than 24.1 and a 95% CI that included 24.0. τ is circadian period.)
- (Step-down primary endpoint) The Clinical Response rate corresponding to individuals who had both entrainment of the aMT6s rhythm and a score of ≥3 on the Non-24 Clinical Response Scale (N24CRS). N24CRS is a 4-item scale that includes the Lower Quartile-nighttime Total Sleep Time (LQ-nTST), Upper Quartile-daytime Total Sleep Duration (UQ-dTSD), Midpoint of Sleep Time (MoST), and Clinician Global Impression-Change (CGI-C) assessments.

The proportion of Non-24 patients who were entrained after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of Non-24 patients who were entrained after placebo treatment (% difference = 17.4; p = 0.0171). The proportion of patients who were entrained (aMT6s) and had a clinical response rate (N24CRS) ≥3 after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of patients who were entrained and had a clinical response rate ≥3 after placebo treatment (% difference = 23.7; p = 0.0028).

Study 3203

The primary efficacy endpoint was the proportion of non-entrainment of the circadian melatonin rhythm as measured by urinary aMT6s. The proportion of Non-24 patients who became non-entrained to a 24-hour day after randomization to tasimelteon was statistically significantly less than the proportion of Non-24 patients who became non-entrained after randomization to placebo treatment (% difference = -70.0; p = 0.0026).

1.2 Summary of Important Events Related to Statistical Analysis in Study 3201

In the original protocol, the primary endpoint proposed by the sponsor was nTST and the proposed sample size was 160 patients, based on the postulated mean treatment difference of 39 minutes and standard deviation of 66 minutes. In Amendment 6 submitted to the Agency, the sample size was changed from 160 to 100 patients, based on the new postulated mean treatment difference of 30 minutes and standard deviation of 45 minutes. In Amendment 9, the primary endpoint was changed to entrainment and the sample size was reduced to 84 patients. At the time of Amendment 9 (May 21, 2012), 95% of the patients were randomized and 56% of the patients completed the study. Amendment 11 was dated on December 11, 2012 and the trial data was unblinded on December 12, 2012. It is unclear to the Agency how much these changes might have impacted the trial results.

1.3 Summary of Statistical Issues

Study 3201

There are three main statistical issues for this study:

Primary Efficacy endpoint(s)

Sponsor's primary efficacy analyses were based on a melatonin-based biomarker endpoint "entrainment". However, throughout the development program for tasimelteon, the Agency repeatedly disagrees that this biomarker is adequate to assess the efficacy for this indication and

requires clinically meaningful primary endpoint(s). No agreement was reached between the Agency and the sponsor regarding the primary endpoint(s). The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the following clinical endpoints: LQ-nTST, UQ-dTSD, MoST, CGIC, nTST and dTSD.

Analysis Populations

For Study 3201, 84 patients were randomized (n=84). The ITT Population defined by the sponsor included all patients randomized into the study that had τ calculated post-randomization (thereafter referred as Sponsor ITT, n=78). The Analysis Population defined by the sponsor included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization (thereafter referred as Sponsor Analysis Population, n=72). The sponsor's ITT and Analysis Population were selected after the randomization and they are non-randomized subsets. Six patients in the Randomized Population without τ calculated post-randomization were excluded from the Sponsor ITT and six patients in the Sponsor ITT were excluded from the Sponsor Analysis Population due to less than 70% of 1 circadian cycle of nTST data reported during screening and/or post-randomization. Thus, 12 patients in total were excluded from the 84 patients randomized. In sponsor's efficacy analyses, the Sponsor ITT population was utilized for all circadian rhythm related outcomes and the Analysis Population was used for analyses of all other endpoints including the step-down primary and all other efficacy analyses. However, the 84 randomized patients all took study medication and had at least one baseline and postbaseline assessment. Based on the intent-to-treat principal, all the 84 patients should be included in the ITT population (n=84). The ITT population is a randomized population. The efficacy analyses conducted by this reviewer are based on ITT population. In Section 3.2.3.4, this reviewer will present information regarding the 12 patients excluded from the efficacy analyses by the sponsor and explain why these 12 patients should be included in the efficacy analyses.

Analysis Methods

In sponsor's efficacy analyses for the clinical endpoints, ANCOVA model was used which includes baseline value as a covariate and treatment group and pooled sites as factors. However, due to small sample size, non-normal distribution of the data, and some heterogeneity in variances, this reviewer thinks permutation ANCOVA is more appropriate than ANCOVA for analyzing the clinical endpoints. In addition, in sponsor's ANCOVA analysis for Study 3201, study site was included as a factor, but the Study Report shows that the randomization was not stratified by study site. Normally, if the randomization isn't stratified by study site, in order to comply with the trial design, the site is not necessarily included in the analysis model. Therefore, this reviewer thinks permutation ANCOVA without site is more appropriate than ANCOVA with site.

Study 3203

For this study, 20 randomized patients all received the study medication and completed the study, thus all 20 patients were included in the ITT population. The issues related to primary endpoint and analysis method are similar to Study 3201, but please note that for Study 3203, the randomization was not stratified by site and site was not a factor in sponsor's ANCOVA analysis.

1.4 Reviewer's Efficacy Analyses

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The analyses presented below were conducted by this reviewer and they are post-hoc exploratory analyses without multiplicity adjustment.

In order to visually examine the distribution of the clinical endpoints, this reviewer generated histograms for each of the clinical endpoints in both studies. Please refer to Section 3.2.3.3 for details.

For Study 3201, sponsor's efficacy analyses for clinical endpoints were based on Analysis Population. The Analysis Population included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization. Based on this definition, 12 patients in the Randomized Population were excluded from the efficacy analysis for clinical endpoints. However, among these 12 patients, 10 patients have more than 50% of 1 circadian cycle of nTST at baseline and 6 patients have more than 50% of 1 circadian cycle of nTST at postbaseline. It seems that the data are not extremely sparse for these 12 patients and this reviewer thinks they should be included in the efficacy analysis. It appears that the inclusion of the 12 patients neither consistently strengthen nor consistently weaken the treatment effect for different clinical endpoints. Please refer to Section 3.2.3.4 for details.

This reviewer conducted ANCOVA analysis on ITT for clinical endpoints. In sponsor's ANCOVA analysis for Study 3201, baseline was included as a covariate if applicable and the pooled sites as a factor (variable name: SITEGR3). However, since the randomization wasn't stratified by site, this reviewer thinks it isn't necessary to include site as a factor in the analysis model. Based on the results of ANCOVA analysis without site, it seems that for both studies the nominal p-values were statistically significant or marginally significant for LQ-nTST, UQ-dTSD, MoST, CGIC and dTSD. The nominal p-value for nTST wasn't statistically significant in either of the studies. Please refer to Section 3.2.3.5 for details.

Because of small sample size, non-normal distribution of the clinical endpoints, and some heterogeneity in the variances of the clinical endpoints, this reviewer thinks permutation ANCOVA would be more appropriate than ANCOVA. It appears that the results of permutation ANCOVA without site is fairly close to those of ANCOVA without site. For LQ-nTST, UQ-dTSD, MoST, CGIC, and dTSD, the nominal p-values are statistically significant or marginally significant for both studies. The p-value for nTST isn't statistically significant for either of the studies. Please refer to Section 3.2.3.6 for details.

This reviewer performed subgroup analysis by sex, race, age group (<50 and >=50) and country (Germany and US) for Study 3201. It seems that the point estimates of treatment effects are all in the right direction for sex, age group and country for all clinical endpoints. For race, the point estimates of treatment effect for LQ-nTST, MoST and CGIC aren't in the right direction for non-white patients. However, this doesn't raise concerns since there are only 14 non-white patients in this study. The subgroup analysis for study 3203 wasn't performed due to small sample size (n=20). Please refer to Section 4.1 for details.

2 INTRODUCTION

2.1 Overview

Non-24 is a severe, chronic, circadian rhythm disorder characterized by the inability to entrain (synchronize) the master body clock to the 24-hour day. Patients with Non-24 have prolonged periods of misalignment of circadian rhythms, including the timing of melatonin and cortisol secretion and the sleep-wake cycle, which are associated with significant impairments in social and occupational functioning, or marked subjective distress.

The majority of reported cases of Non-24 occur in blind patients with no conscious perception of light. As a result of light information failing to reach the suprachiasmatic nuclei (SCN) to synchronize the clock and its outputs, the pacemaker may revert to its endogenous non-24-hour period. Consequently, the timing of physiology and behavior that is controlled by the circadian system (e.g. the timing of melatonin and cortisol production, the core body temperature rhythm, metabolic processes, the sleep-wake cycle, and alertness and performance patterns) becomes desynchronized from the 24-hour day, which has serious consequences on the daily functioning of the patient. The estimated prevalence of Non-24 in the totally blind is approximately 100,000 individuals in the United States. Studies suggest that 50-70% of totally blind individuals with no light perception have Non-24.

Currently, there is no approved treatment for Non-24. The sponsor thinks tasimelteon is a circadian regulator that resets the master body clock in the suprachiasmatic nucleus (SCN). Tasimelteon is believed to be a Dual Melatonin Receptor Agonist (DMRA) with selective agonist activity at the MT_1 and MT_2 melatonin receptors.

The development of tasimelteon formerly referred to as VEC-162 and BMS-214778 was initiated by Bristol Myers Squibb (IND Submission #1 December 17, 1997). Vanda Pharmaceuticals Inc. (Vanda) licensed the tasimelteon product in 2004 and was granted Orphan drug designation for the treatment of Non-24-Hour Disorder in blind individuals with no light perception on January 9, 2010.

This submission includes two pivotal efficacy studies.

Study 3201 was a multicenter, randomized, double-masked, placebo-controlled, parallel study designed to evaluate the efficacy and safety of 20 mg of tasimelteon versus placebo in patients suffering from Non-24. Eighty-four patients (tasimelteon 42; placebo 42) were randomized to receive tasimelteon (20 mg/day) or placebo. This study was conducted at investigative sites in the US and Germany. The study began with a Pre-Randomization Phase (~5-6 weeks) and was followed by either a Randomization Phase (~26 weeks) or an Open-Label Extension Phase (~26 weeks).

Study 3203 was a multicenter, randomized withdrawal, double-masked, placebo-controlled, parallel group study designed to evaluate the long-term maintenance effect and safety of 20 mg of tasimelteon versus placebo in patients with Non-24. Every patient who enrolled in Study 3203 had previously been screened in 3201. Patients who met the inclusion criteria and who had previously participated in, or were screened for Study 3201, were eligible to participate. The study had 2 phases: a Pre-Randomization Phase (consisting of an Open-label tasimelteon Run-in Phase [~6 weeks] and a τ Estimation Phase [~ 6 weeks]), and a Randomized Withdrawal Phase (~8 weeks).

For Study 3201, the first patient was enrolled on August 25, 2010 and the last patient completed on October 29, 2012. For Study 3203, the first patient was enrolled on September 15, 2011 and the last patient completed on November 28, 2012. According to the sponsor, the date of data unblinding was December 12, 2012 for Study 3201 and January 16, 2013 for Study 3203. The Statistical Analysis Plan for Study 3201 and Study 3203 were singed-off on December 11, 2012 and December 28, 2012, respectively.

2.2 Data Sources

The sponsor's electronic submission was stored in the directory of \\Cdsesub1\evsprod\NDA205677 of the center's electronic document room.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the clinical endpoints. The analyses conducted by this reviewer were produced from raw data.

3.2 Evaluation of Efficacy

Reviewer's Note: Section 3.2.1 and Section 3.2.2 present the study design, statistical analysis plan and efficacy results excerpted from sponsor's protocols and Clinical Study Reports for Study 3201 and Study 3203, respectively. Please note that there is no agreement between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods.

3.2.1 PROTOCOL VP-VEC-162-3201 (SET STUDY)

3.2.1.1 Study Objectives

The primary objective of this study was:

- To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of entrainment.
- (a step-down objective): To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of patients with a clinical response. Clinical response was defined as the coincident demonstration of entrainment of the 6-sulfatoxymelatonin (aMT6s) rhythm and a score of ≥3 on the Non-24 Clinical Response Scale (N24CRS).

The key secondary objectives were:

- To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of responders with a combined sleep/wake response for nighttime sleep duration and daytime sleep.
- To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of entrainment as assessed by urinary cortisol.

Additional secondary objectives of this study were:

- Subtype I: To determine the efficacy of tasimelteon in patients with Non-24, as measured by the proportion of subtype I responders, defined as an individual who was both entrained and had a significant improvement from screening in Lower Quartile – nighttime Total Sleep Time (LQ-nTST);
- Subtype II: To determine the efficacy of tasimelteon in patients with Non-24, as measured by the proportion of subtype II responders, defined as an individual who was both entrained and had a significant improvement from screening in Upper Quartile – daytime Total Sleep Duration (UQ-dTSD);
- Subtype III: To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of subtype III responders, defined as an individual who was both entrained and had a significant improvement from screening in Midpoint of Sleep Timing (MoST);
- Subtype IV: To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of subtype IV responders, defined as an individual who was both entrained and had a significant improvement from screening in Clinical Global Impression-Change (CGI-C);
- To determine the association between treatment response, as measured by entrainment of aMT6s circadian rhythms, and the baseline urinary aMT6s excretion rate in tasimelteon-treated patients with Non-24;
- To determine the efficacy of tasimelteon in improving subjective nighttime total sleep time (nTST) in patients with Non-24, as assessed by the change from screening in the average of LQ-nTST;
- To determine the efficacy of tasimelteon in reducing subjective daytime total sleep duration (dTSD) in patients with Non-24, as assessed by the change from screening in the average of UQ-dTSD;
- To determine the efficacy of tasimelteon in patients with Non-24 as measured by the change from the screening in MoST;
- To determine the efficacy of tasimelteon to treat Non-24, as assessed by the CGI-C;
- To determine the efficacy of tasimelteon in patients with Non-24 as measured by the proportion of entrainment as assessed by urinary analytes under circadian control.

3.2.1.2 Study Design

This was a multicenter, randomized, double-masked, placebo-controlled, parallel study designed to evaluate the efficacy and safety of 20 mg of tasimelteon versus placebo in patients suffering from Non-24. Eight-four patients (tasimelteon 42; placebo 42) were randomized in a ratio of 1:1 to receive tasimelteon (20 mg/day) or placebo. This study was conducted at investigative sites in the US and Germany. The study began with a Pre-Randomization Phase (~5-6 weeks) and was followed by either a Randomization Phase (~26 weeks) or an Open-Label Extension Phase (~26 weeks).

The Pre-Randomization Phase comprised a screening visit, a circadian period (τ) estimation segment, and a variable-length in-phase transition segment.

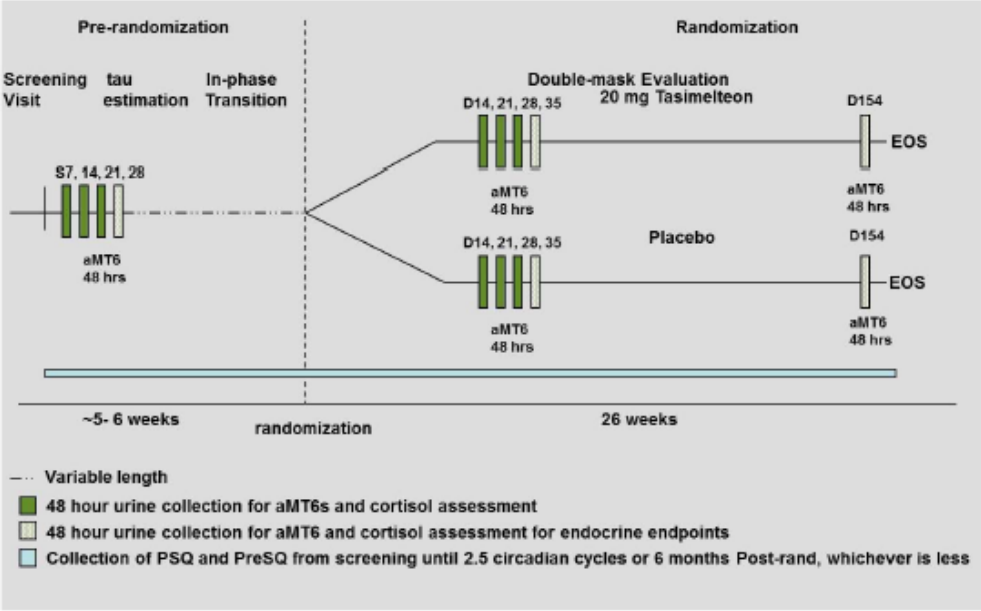
After the Pre-Randomization Phase, patients who met all entry criteria for the study entered the Randomization Phase. Eligible patients at US sites could participate in the Randomization Phase or the Open-Label Extension Phase; eligible patients at sites in Germany could participate in the Randomization Phase followed by the Open-Label Extension Phase after completing the washout segment or could directly enter the Open-Label Extension Phase. Per Protocol Amendment 6 for US sites, the 2-week washout segment was removed to directly enroll patients into Study VP-VEC-162-3203 (a randomized withdrawal study) or Study VP-VEC-162-3204 (a safety study), without interruption of treatment.

The Randomization Phase comprised the double-masked evaluation segment and a 2-week washout segment (for German sites and US patients who completed the evaluation segment prior to implementation of Protocol Amendment 6). During the Randomization Phase, patients were asked to take either 20 mg tasimelteon or placebo approximately 1 hour prior to their target bedtime for 26 weeks in a double-masked fashion. Patients reported their nighttime sleep parameters for 2.5 circadian cycles (one circadian cycle = the number of calendar days for the non-24-hour circadian rhythms to complete a 24-hour cycle; e.g., 48 days for a $\tau = 24.5$ hours) or 6 months, whichever was less. Patients also completed 48-hour urine collections for aMT6s assessment, cortisol, and other analyte assessments beginning on Days D14, D21, D28, D35, and D154. During the washout segment, patients were treated for 2 weeks with placebo in a single-masked fashion.

Patients who completed the Randomization Phase of the study were given the opportunity to participate in the optional Open-Label Extension Phase (for German sites). Additionally, patients who had a $\tau > 24.0$ and met all entry criteria but were ineligible for the Randomization Phase due to their τ were given the opportunity to participate in the Open-Label Extension Phase. During this phase, patients took open-label 20 mg tasimelteon for 26 weeks. The purpose of the Open-Label Extension Phase was to explore the long-term safety of tasimelteon in patients with Non-24 over 26 weeks.

Figure 1 and Figure 2 provide schematics of the study design for sites in US and Germany.

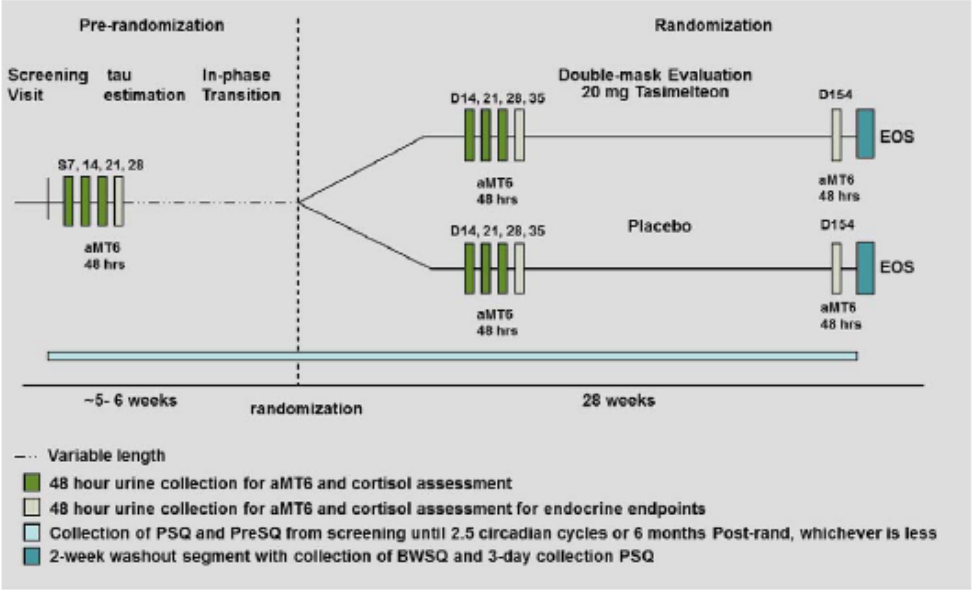
Figure 1: Study Design (US)



aMT6s = 6-sulfatoxymelatonin; EOS = end-of-study; Post-rand = post-randomization; PreSQ = Pre-Sleep Questionnaire; PSQ = Post-Sleep Questionnaire.
Source: Study Protocol (Appendix 16.1.1)

Source: Figure 1 of sponsor's Clinical Study Report

Figure 2: Study Design (Germany)



aMT6s = 6-sulfatoxymelatonin; BWSQ = Benzodiazepine Withdrawal Symptom Questionnaire; EOS = end-of-study; Post-rand = post-randomization; PreSQ = Pre-Sleep Questionnaire; PSQ = Post Sleep Questionnaire.
Source: Study Protocol (Appendix 16.1.1)

Source: Figure 2 of sponsor's Clinical Study Report

3.2.1.3 Efficacy Measures

Efficacy measurements included assessments of aMT6s, cortisol, other circadian analytes, Post Sleep Questionnaire (PSQ), Pre-Sleep Questionnaire (PreSQ), and Clinical Global Impression of Change (CGI-C).

The primary efficacy endpoints were the following:

- The entrainment of the circadian melatonin rhythm as measured by urinary aMT6s (Entrainment is a melatonin-based biomarker and is defined as having a post-baseline τ value less than 24.1 and a 95% CI that included 24.0. τ is circadian period.)
- (Step-down primary endpoint) The Clinical Response rate corresponding to individuals who had both entrainment of the aMT6s rhythm and a score of $\geq$3 on the Non-24 Clinical Response Scale (N24CRS).

N24CRS is a 4-item scale that includes the LQ-nTST, UQ-dTSD, MoST, and CGI-C assessments (Table 1). Each assessment on the scale is scored as a 1 or 0 depending on whether the pre-specified threshold was achieved or not. The score for each assessment was summarized with a range of 0 to 4.

Table 1: Non-24 Scale of Clinical Response (N24CRS)

Assessment	Threshold of Response
LQ-nTST	>45 minutes increase in average nighttime sleep duration
UQ-dTSD	>45 minutes decrease in average daytime sleep duration
MoST	>30 minutes increase and a standard deviation <2 hours during double-masked phase
CGI-C	<2.0 from the average of D112 and D183 compared to baseline

Source: Table 7 of sponsor's Clinical Study Report

The key secondary efficacy endpoints were the following:

- The proportion of responders with a combined sleep/wake response for nighttime sleep duration and daytime sleep duration defined as increase of 90 minutes or greater in LQ-nTST and decrease of 90 minutes or greater in UQ-dTSD.
- The entrainment of the circadian rhythm as measured by urinary cortisol.

The additional secondary efficacy outcomes were the following:

- The subtype I response rate (sleep time subtype)
- The subtype II response rate (daytime sleep subtype)
- The subtype III response rate (MoST subtype)
- The subtype IV response rate (CGI-C subtype)
- Treatment response association with aMT6s excretion rate
- The average of LQ-nTST
- The average of UQ-dTSD
- The average of MoST
- The CGI-C
- The entrainment of circadian analytes.

3.2.1.4 Statistical Analysis Plan

Below is a summary of sponsor's Statistical Analysis Plan.

The following analysis populations were defined for this study:

- Intent-to-Treat (ITT) Population: The ITT Population included all patients randomized into the study that had τ calculated post-randomization.
- Analysis Population: The Analysis Population included all patients in the ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization. The number of non-missing days of nTST data during each of screening and post-randomization had to be at least equal to the number of days that made up 70% of 1 circadian cycle.
- Safety Population: The Safety Population included all patients randomized into the study who received at least 1 dose of study drug.

The ITT population was utilized for all circadian rhythm related outcomes: the primary endpoint of entrainment, the cortisol entrainment endpoint, and the aMT6s baseline secretion rate analyses. The Analysis Population was used for analysis of all other endpoints including the step-down primary and all other efficacy analyses.

The primary efficacy endpoint was defined as the proportion of the number of Non-24 patients who were entrained after placebo or tasimelteon treatment during the Randomization Phase of Study 3201 only. Hence, the primary null hypothesis of the primary efficacy endpoint was that no difference exists in the proportion of people with entrained circadian rhythms between patients receiving tasimelteon and patients receiving placebo. Barnard's Exact Test was used to test the null hypothesis in the ITT Population. The Fisher's Exact Test was used as the sensitivity analysis method.

If the primary null hypothesis was rejected at an alpha level of 0.05, then the step-down primary null hypothesis was to be tested at an alpha level of 0.05 to assess the efficacy of tasimelteon versus placebo as measured by the Clinical Response Rate. The Clinical Response Rate was summarized and analyzed in the same manner of the primary endpoint.

Sensitivity analyses were conducted to verify the results of the step-down endpoint (Clinical Response). The first analysis was conducted for this step-down endpoint but with threshold scores of ≥2 for the N24CRS. The second and the third were conducted for individuals (regardless of entrainment status) with a N24CRS score of ≥3 and ≥2, respectively. The last one was conducted for responders who were entrained in the Randomization Phase of Study 3201 only that also had an N24CRS score of ≥3. This differs from the step-down primary endpoint which defined entrainment as occurring in the Randomization Phase of Study 3201 or the Run-in Phase of Study 3203 for individuals who were originally treated in the tasimelteon arm of Study 3201. The clinical measurements in the N24CRS for all analyses were only derived from the Randomization Phase of Study 3201.

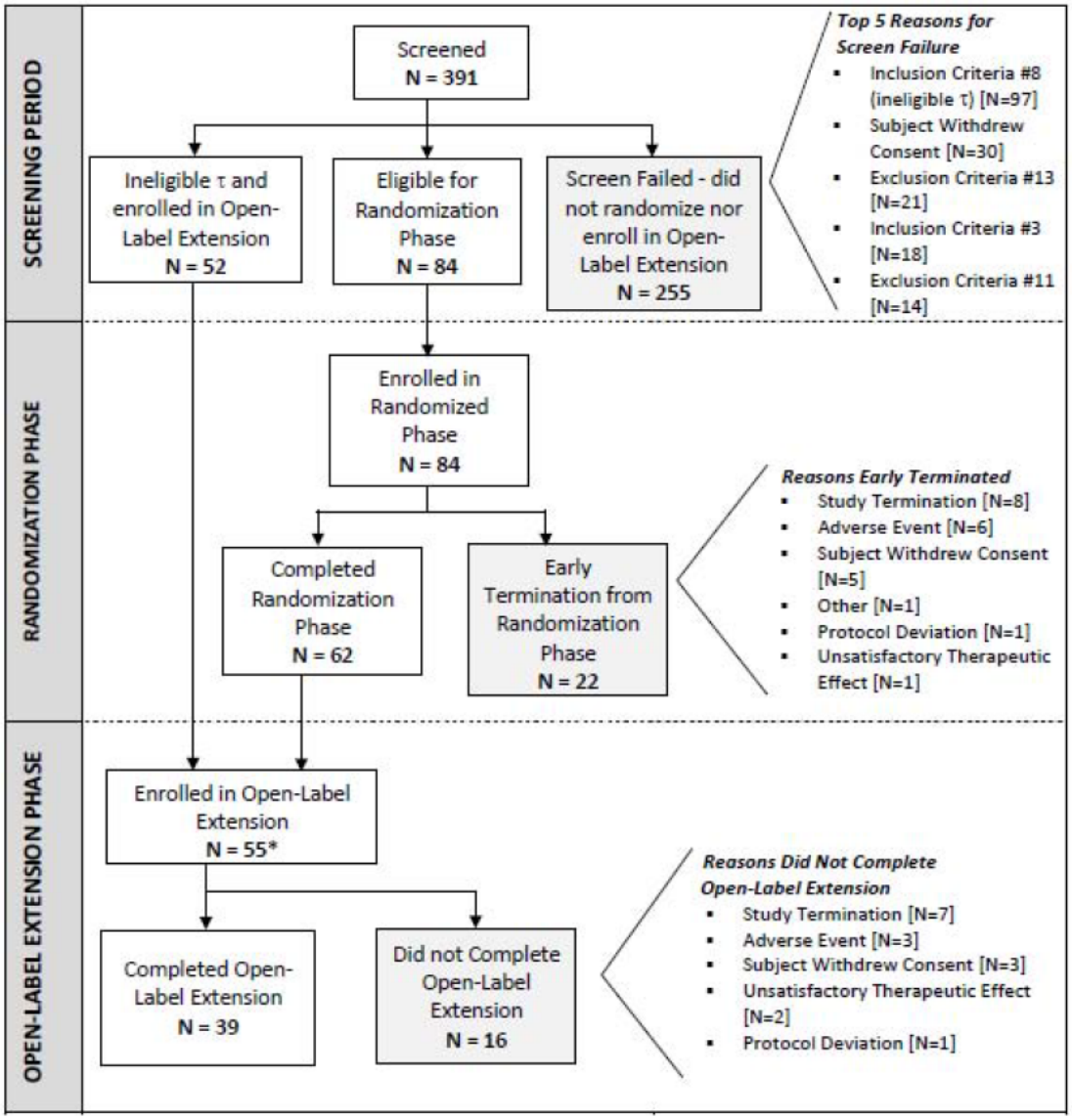
For key secondary efficacy endpoints, the cortisol entrainment rate was summarized and analyzed in the same manner of the primary endpoint in the ITT Population. In addition, the sleep/wake response endpoint was summarized and analyzed in the same manner of the primary endpoint in the Analysis Population.

3.2.1.5 Patient Disposition, Demographic and Baseline Characteristics

Disposition of Patients

Figure 3 shows the patient disposition of all patients.

Figure 3: Flow Diagram of Patients Disposition (All Patients)



* 52 patients entered the Open-Label Extension Phase from screening and 3 patients rolled over after completing the Randomization Phase.

Source: Post-text Tables 14.1.1, 14.1.2, 14.1.4, and 14.1.5

Source: Figure 3 of sponsor's Clinical Study Report

Of the 391 patients who were screened for this study, 255 patients failed screening and were not enrolled in the Randomization Phase or the Open-Label Extension Phase, 84 patients were enrolled in the Randomization Phase, and 52 patients failed screening and were enrolled in the Open-Label Extension Phase.

An equal number of patients were randomized into each treatment group of the Randomization Phase: 42 patients in the placebo group and 42 patients in the tasimelteon group. Of these 84 patients who were randomized to study drug, 62 (73.8%) patients completed the phase and 22 (26.2%) patients discontinued early. The reasons for discontinuation during the Randomization Phase were comparable between placebo and tasimelteon treatment groups.

Datasets Analyzed

Data sets analyzed for the Randomization Phase and the Open-Label Extension Phase are presented in Table 2. For the Randomization Phase, all 84 patients received at least one dose of study drug and were included in the Safety Population. The ITT Population included 78 randomized patients who had a τ value calculated post-randomization. The Analysis Population included 72 patients from the ITT Population who had at least 70% of one circadian cycle of nTST data reported during screening and post-randomization. For the Open-Label Extension Phase, 54 patients received at least one dose of study drug and were included in the Safety Population.

Table 2: Summary of Analysis Datasets

	Placebo (N=42) n (%)	**Tasimelteon 20 mg (N=42) n (%)**	**Total (N=84) n (%)**
Randomization phase[1]			
Randomized	42 (100.0)	42 (100.0)	84 (100.0)
Did not receive study drug	0 (0.0)	0 (0.0)	0 (0.0)
Safety population	42 (100.0)	42 (100.0)	84 (100.0)
Intent-to-treat population	38 (90.5)	40 (95.2)	78 (92.9)
Analysis population	34 (81.0)	38 (90.5)	72 (85.7)
Inadequate sleep assessment during screening and/or randomization phase	4 (9.5)	2 (4.8)	6 (7.1)
Open-label extension phase[2]			
Open-label extension safety population	--	54 (98.2)	--
Enrolled in open-label extension phase but did not take any study drug	--	1 (1.8)	--

[1] Percentages were based on the total number of patients randomized (N).
[2] Percentages were based on the number of patients enrolled in the Open-Label Extension Phase.
Source: Post-text Tables 14.1.6 and 14.1.2

Source: Table 12 of sponsor's Clinical Study Report

Demographics and Baseline Characteristics

Table 3 summarizes demographics and baseline characteristics for all randomized patients. There were 49 males and 35 females in the study. The mean age was 50.7 years with a range of 23 to 74 years. The majority of patients were White (83.3%) and not Hispanic or Latino (96.4%). Overall, the mean LQ-nTST, UQ-dTSD, nTST, and dTSD at screening were 3.2 h, 2.4 h, 5.3 h, and 0.9 h, respectively. The mean τ values as measured by urinary aMT6s and cortisol were 24.47 h and 24.45 h, respectively. It seems that demographic and baseline characteristics were comparable between the groups.

Table 3: Demographics and Baseline Characteristics – All Randomized Patients

	Placebo (N=42)	**Tasimelteon 20 mg (N=42)**	**Total (N=84)**
Age at screening (years)			
n	42	42	84
Mean (SD)	50.7 (13.15)	50.8 (12.63)	50.7 (12.82)
Age category (n, %)			
18 to 40 years	9 (21.4)	10 (23.8)	19 (22.6)
41 to 65 years	28 (66.7)	29 (69.0)	57 (67.9)
>65 years	5 (11.9)	3 (7.1)	8 (9.5)
Gender (n, %)			
Male	25 (59.5)	24 (57.1)	49 (58.3)
Female	17 (40.5)	18 (42.9)	35 (41.7)
Race			
American Indian or Alaska native	1 (2.4)	0 (0.0)	1 (1.2)
Asian	0 (0.0)	1 (2.4)	1 (1.2)
Black or African or African American	6 (14.3)	4 (9.5)	10 (11.9)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)
White	34 (81.0)	36 (85.7)	70 (83.3)
Other	1 (2.4)	1 (2.4)	2 (2.4)
Ethnicity (n, %)			
Hispanic or Latino	1 (2.4)	2 (4.8)	3 (3.6)
Not Hispanic or Latino	41 (97.6)	40 (95.2)	81 (96.4)
Weight at screening (kg)			
n	42	42	84
Mean (SD)	78.58 (17.216)	80.19 (16.311)	79.39 (16.688)

	Placebo (N=42)	Tasimelteon 20 mg (N=42)	Total (N=84)
Body mass index at screening (kg/m^2)			
n	42	42	84
Mean (SD)	27.73 (3.930)	28.16 (4.055)	27.95 (3.975)
Average of LQ-nTST (hours)			
n	42	42	84
Mean (SD)	3.246 (1.6071)	3.245 (1.1902)	3.246 (1.4055)
Median	2.982	3.383	3.089
Average of UQ-dTSD (hours)			
n	42	42	84
Mean (SD)	2.533 (1.7084)	2.290 (1.6558)	2.411 (1.6766)
Median	2.095	2.007	2.074
Average nTST (hours)			
n	42	42	84
Mean (SD)	5.416 (1.2689)	5.243 (1.0436)	5.330 (1.1580)
Median	5.154	5.160	5.154
Average dTSD (hours)			
n	42	42	84
Mean (SD)	0.974 (0.7415)	0.859 (0.7740)	0.917 (0.7555)
Median	0.769	0.648	0.715
τ (aMT6s) (hours)			
n	42	42	84
Mean (SD)	24.45 (0.172)	24.49 (0.158)	24.47 (0.165)
Median	24.41	24.49	24.45
τ (cortisol) (hours)			
n	42	42	84
Mean (SD)	24.43 (0.325)	24.48 (0.254)	24.45 (0.291)
Median	24.42	24.47	24.46

aMT6s = 6-sulfatoxymelatonin; dTSD = daytime total sleep duration; LQ-nTST = lower quartile of subjective nighttime total sleep time; nTST = nighttime total sleep time; SD = standard deviation; UQ-dTSD = upper quartile of subjective daytime total sleep duration; SD = standard deviation.
Source: Post-text Table 14.1.8.1

Source: Table 13 of sponsor's Clinical Study Report

3.2.1.6 Sponsor's Primary Efficacy Results

The ITT population was utilized for all circadian rhythm related outcomes: the primary endpoint of entrainment, the cortisol entrainment endpoint, and the aMT6s baseline secretion rate

analyses. The Analysis Population was used for analysis of all other endpoints including the step-down primary and all other efficacy analyses. Table 4 presents a summary of the results of the primary efficacy endpoints.

Table 4: Summary of Primary Efficacy Endpoints (ITT and Analysis Population)

Category **Statistic**	**Placebo** **n (%)**	**Tasimelteon** **20 mg** **n (%)**	**P-value[1]**	**P-value[2]**
Primary analysis	(N=38)	(N=40)		
Entrainment rate (aMT6s)[3]	1 (2.6)	8 (20.0)	0.0171	0.0291
Primary step-down analysis	(N=34)	(N=38)		
Entrainment and N24CRS ≥3[4,5]	0 (0.0)	9 (23.7)	0.0028	0.0025
Sensitivity analyses	(N=34)	(N=38)		
Entrainment in Study 3201 and N24CRS ≥3[5]	0 (0.0)	5 (13.2)	0.0286	0.0558
Entrainment and N24CRS ≥2[4,5]	0 (0.0)	11 (28.9)	0.0006	0.0005
N24CRS ≥3[5]	1 (2.9)	11 (28.9)	0.0031	0.0036
N24CRS ≥2[5]	7 (20.6)	22 (57.9)	0.0014	0.0017

[1] P-value was based on Barnard's Exact Test, two-sided.
[2] P-value was based on Fisher's Exact Test, two-sided.
[3] Entrainment was defined as having a post-baseline τ value <24.1 and a 95% CI that included 24.0. Results were based on the ITT Population from Study 3201 only.
[4] Entrainment was defined as having a post-baseline τ value <24.1 and a 95% CI that included 24.0. For patients randomized to tasimelteon 20 mg and who participated in the screening phase of Study 3203, the screening τ from Study 3203 was used if the patient did not become entrained in Study 3201 but did become entrained during the screening phase of Study 3203. Results were based on the Analysis Population.
[5] Non-24 Clinical Response Scale was a 4-item scale that included LQ-nTST, UQ-dTSD, MoST, and CGI-C assessments. Each assessment was scored as 1 or 0 depending on whether the subtype responder was achieved. The range of N24CRS was 0 to 4. Results were based on the Analysis Population.
aMT6s = 6-sulfatoxymelatonin; ITT = intent-to-treat; LQ-nTST = lower quartile of subjective nighttime total sleep time; MoST = midpoint of sleep timing; N24CRS = Non-24 Clinical Response Scale; Study 3201 = Study VP-VEC-162-3201; Study 3203 = Study VP-VEC-162-3203; UQ-dTSD = upper quartile of subjective daytime total sleep duration.
Source: Post-text Tables 14.2.1, 14.2.2.1, 14.2.2.2, 14.2.2.3, 14.2.3.1, 14.2.3.2

Source: Table 14 of sponsor's Clinical Study Report

Entrainment was defined as having a post-baseline τ value less than 24.1 and a 95% CI that included 24.0. The proportion of Non-24 patients who were entrained after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of Non-24 patients who were entrained after placebo treatment (% difference = 17.4; p = 0.0171).

The Clinical Response Rate was defined as the coincident demonstration of entrainment of the aMT6s rhythm and a score of ≥3 on the N24CRS. The proportion of patients who were entrained (aMT6s) and had a clinical response rate (N24CRS) ≥3 after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of patients who were entrained and had a clinical response rate ≥3 after placebo treatment (% difference = 23.7; p = 0.0028).

3.2.1.7 Sponsor's Secondary Efficacy Results

Table 5 presents a summary of the results of the secondary efficacy endpoints during the Randomization Phase. Please note that all p-values in this table are nominal p-values.

Table 5: Summary of Secondary Efficacy Endpoints (ITT and Analysis Population)

Category Statistic	Placebo	Tasimelteon 20 mg	Treatment Difference (Tasimelteon minus Placebo)	P-value
LQ-nTST and UQ-dTSD ≥45 min, n (%)[1,2]	3 (8.8)	12 (31.6)	22.8	0.0177
LQ-nTST and UQ-dTSD ≥90 min, n (%)[1,2]	1 (2.9)	5 (13.2)	10.2	0.1312
Entrainment rate (cortisol), n (%)[1,3,4]	1 (2.6)	7 (17.5)	14.9	0.0313
Entrained and ≥45 minute improvement in LQ-nTST, n (%)[1,2,3]	0 (0.0)	10 (26.3)	26.3	0.0013
LQ-nTST, LS mean hours (LS mean minutes)[5]	0.28 (17.08)	0.95 (56.80)	0.66 (39.71)	0.0055
Entrained and ≥45 minute improvement in UQ-dTSD, n (%)[1,2,3]	0 (0.0)	11 (28.9)	28.9	0.0006
UQ-dTSD, LS mean hours (LS mean minutes)[5]	-0.30 (-17.87)	-0.77 (46.48)	-0.48 (-28.61)	0.0050
Entrained and ≥30 min improvement in MoST and SD ≤2.0 hours, n (%)[1,2,3]	0 (0.0)	8 (21.1)	21.1	0.0046
MoST, LS mean hours (LS mean minutes)[5]	0.24 (14.48)	0.58 (35.00)	0.34 (20.52)	0.0123
Entrained and ≤2.0 in average CGI-C, n (%)[1,2,3]	0 (0.0)	7 (19.4)	19.4	0.0078
CGI-C (LS mean)[6]	3.4	2.6	-0.8	0.0093

[1] P-value was based on Barnard's Exact Test, two-sided.
[2] Results for LQ-nTST, UQ-dTSD, MoST, and CGI-C assessments were based on the Analysis Population where N=34 in the placebo group and N=38 in the tasimelteon group; the LQ-nTST and UQ-dTSD ≥45 minute sleep/wake response was a post-hoc analysis.
[3] Entrainment was defined as having a post-baseline τ value <24.1 and a 95% CI that included 24.0. For patients randomized to tasimelteon 20 mg and who participated in the screening phase of Study 3203, the screening τ from Study 3203 was used if the patient did not become entrained in Study 3201 but did become entrained during the screening phase of Study 3203.
[4] Results were based on the ITT Population where N=38 in the placebo group and N=40 in the tasimelteon group.
[5] P-value was based on analysis of covariance model.
[6] P-value was based on analysis of variance model.
CGI-C = Clinical Global Impression-Change; CI = confidence interval; ITT = intent-to-treat; LQ-nTST = lower quartile of subjective nighttime total sleep time; LS mean = least-squares mean; MoST = midpoint of sleep timing; N24CRS = Non-24 Clinical Response Scale; UQ-dTSD = upper quartile of subjective daytime total sleep duration.
Source: Post-text Tables 14.2.4.1, 14.2.4.3, 14.2.5, 14.2.6.1, 14.2.6.2, 14.2.6.3, 14.2.6.4, 14.2.7.1, 14.2.7.2, 14.2.8.1, 14.2.8.2, 14.2.9.1, 14.2.9.2, 14.2.10.1

Source: Table 15 of sponsor's Clinical Study Report

3.2.2 PROTOCOL VP-VEC-162-3203 (RESET STUDY)

3.2.2.1 Study Objectives

The primary objective of this study was to demonstrate the maintenance of effect of tasimelteon to entrain circadian rhythms in patients with Non-24 as measured by urinary aMT6s.

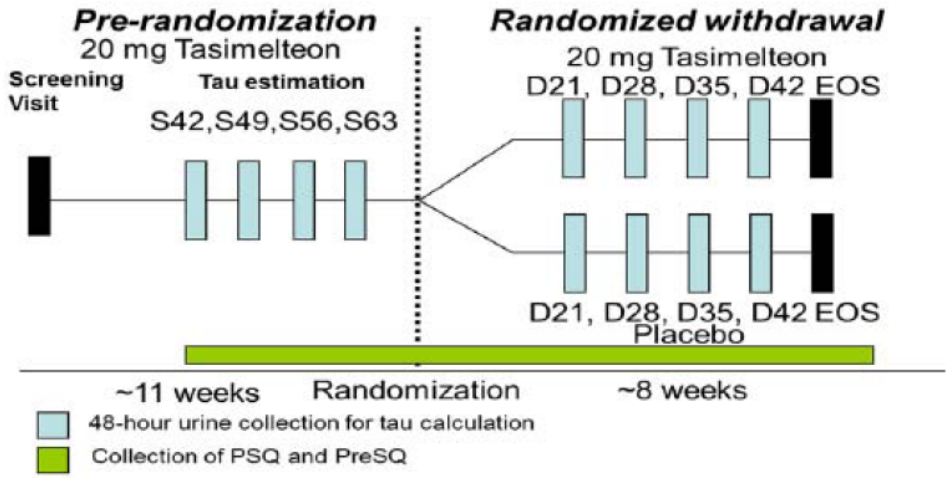
The secondary objectives of this study were the following:

- To demonstrate the maintenance of effect of tasimelteon to treat Non-24 as measured by the time to relapse with relapse defined as a 45 minute or greater decrement in the weekly average subjective nighttime total sleep time (nTST) compared to the Run-in Phase;
- To demonstrate the maintenance of effect of tasimelteon to entrain circadian rhythms in patients with Non-24 as assessed by urinary cortisol;
- To demonstrate the maintenance of effect of tasimelteon in patients with Non-24 as measured by the proportion of patients with non-entrainment and an average of 30 minutes or greater decrement of nTST compared to the Run-in Phase;
- To demonstrate the maintenance of effect of tasimelteon on subjective nTST in patients with Non-24, as assessed by the change from the Run-in Phase in the average nTST;
- To demonstrate the maintenance of effect of tasimelteon on subjective nTST in patients with Non-24 as assessed by the change from the Run-in Phase in the lower quartile of days of nTST (LQ-nTST);
- To demonstrate the maintenance of effect of tasimelteon on subjective daytime total sleep duration (dTSD) in patients with Non-24, as assessed by the change from the Run-in Phase in the average dTSD;
- To demonstrate the maintenance of effect of tasimelteon on subjective dTSD in patients with Non-24, as assessed by the change from the Run-in Phase in the upper quartile of days of dTSD (UQ-dTSD);
- To demonstrate the maintenance of effect of tasimelteon on the midpoint of sleep time (MoST); and
- To assess symptoms of withdrawal after a minimum of 3 months of tasimelteon treatment assessed by the Benzodiazepine Withdrawal Symptom Questionnaire (BWSQ).

3.2.2.2 Study Design

This was a multicenter, randomized withdrawal, double-masked, placebo-controlled, parallel-group study designed to evaluate the long-term maintenance effect and safety of 20 mg of tasimelteon versus placebo in patients with Non-24. Patients who met the entrance criteria and who had previously participated in, or were screened for, Study VP-VEC-162-3201 were eligible to participate. The study had 2 phases: a Pre-Randomization Phase (consisting of an Open-label tasimelteon Run-in Phase [approximately 6 weeks] and a τ Estimation Phase [approximately 6 weeks]), and a Randomized Withdrawal Phase (8 weeks). Twenty patients (tasimelteon 10; placebo 10) were randomized in a ratio of 1:1 to receive tasimelteon (20 mg/day) or placebo during the Randomized Withdrawal Phase. This study was conducted at 18 investigative sites in the US. Figure 4 presents an overview of the study design.

Figure 4: Study Design



EOS = End of study visit; PreSQ = Pre-Sleep Questionnaire; PSQ = Post-Sleep Questionnaire.
Source: Study Protocol (Appendix 16.1.1)

Source: Figure 1 of sponsor's Clinical Study Report

3.2.2.3 Efficacy Measures

Efficacy measurements included assessments of urinary aMT6s, time to relapse, urinary cortisol, nighttime sleep, daytime naps, and timing of sleep.

The primary efficacy endpoint was the following:

- The proportion of non-entrainment of the circadian melatonin rhythm as measured by urinary aMT6s.

The secondary efficacy endpoints were the following:

- Time to relapse with relapse defined as a 45 minute or greater decrement in the weekly average of nTST;
- The proportion of non-entrainment of the circadian rhythm as measured by cortisol;
- Proportion of patients who were non-entrained and had an average of 30 minutes or greater decrement of subjective nTST compared to the Run-in Phase;
- The change in average nTST;
- The change in average LQ-nTST;
- The change in average dTSD;
- The change in average UQ-dTSD; and
- The change in average MoST.

3.2.2.4 Statistical Analysis Plan

The following analysis plan was excerpted from sponsor's Statistical Analysis Plan.

The following analysis populations were defined for this study:

- Intent-to-Treat (ITT) Population: The ITT Population included all randomized patients who had τ calculated post-randomization.

- Safety Population for the Tasimelteon Run-in Phase: The Safety Population for the tasimelteon Run-in Phase included all patients who entered the tasimelteon Run-in Phase and received at least one dose of study drug.

The primary efficacy endpoint was defined as the proportion of patients who become non-entrained to a 24-hour day after randomization to tasimelteon or placebo. Hence, the primary null hypothesis was that no difference existed in the proportion of patients with "non-entrained" circadian rhythms between patients receiving tasimelteon and patients receiving placebo. Barnard's Exact Test was used to test the null hypothesis in the ITT population. The Fisher's Exact Test was used as the sensitivity analysis method.

Circadian cycle time to first relapse event was analyzed via Kaplan-Meier product-limit survival curve estimates and an un-stratified log-rank test for treatment group comparison. Patients who did not meet the criteria for a relapse event (relapse event being defined as a 45 minute decrement in their nTST) during the Randomized Withdrawal Phase were censored at the cycle time of the discontinuation or completion of the study.

The cortisol non-entrainment and the patients who were non-entrained as measured by aMT6s and had an average of 30 minutes or greater decrement of subjective nTST compared to the Run-in Phase, were summarized and analyzed in the same manner as the primary endpoint in the ITT Population.

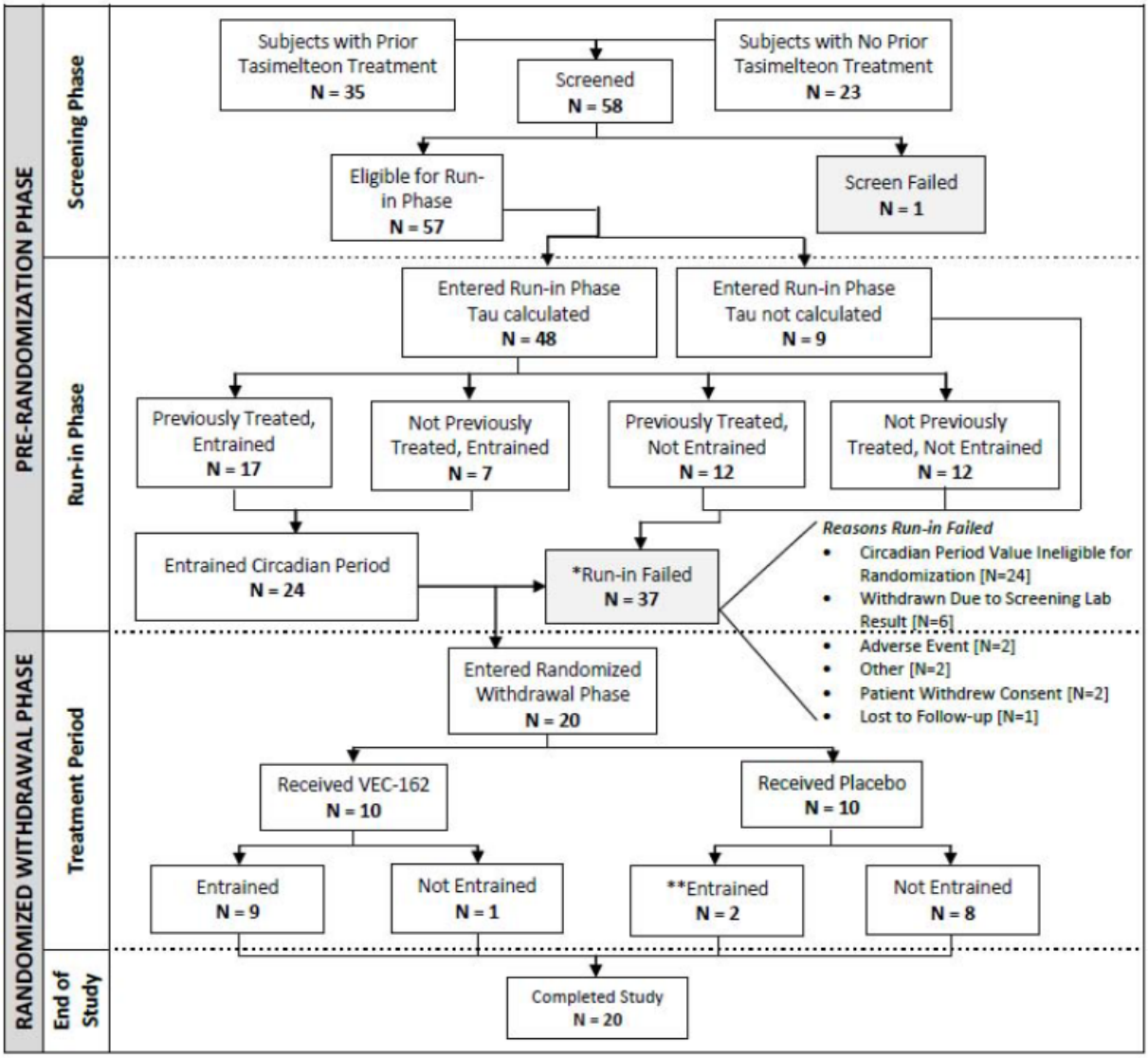
For the analysis of continuous efficacy variables, descriptive statistics were presented by treatment group for all patients in the ITT population. Treatment groups were compared using an analysis of covariance (ANCOVA) model with the terms of treatment group and the corresponding efficacy value in the tasimelteon Run-in Phase as a covariate.

3.2.2.5 Patient Disposition, Demographic and Baseline Characteristics

Patients Disposition

Figure 5 presents the disposition of all patients. The whole Pre-Randomization Phase including the tasimelteon Run-in Phase and the τ Estimation Phase was referred to as the "Run-in Phase." Of the 58 patients who were screened for this study, 1 patient failed screening and was not enrolled in the Run-in Phase. A total of 57 patients received tasimelteon during the Run-in Phase; 37 of those patients failed the Run-in Phase and were not enrolled in the Randomized Withdrawal Phase. An equal number of patients were randomized into each treatment group of the Randomized Withdrawal Phase: 10 patients in the placebo group and 10 patients in the tasimelteon group. All randomized patients completed the Randomized Withdrawal Phase.

Figure 5: Flow Diagram of Patient Disposition (All Patients)



* Four patients were entrained during the Run-in Phase but did not enter the Randomized Withdrawal Phase due to the study already reaching N=20 for randomized patients (N=2), adverse events during the Run-in Phase (N=1), and the patient met the definition for entrainment, but was screen failed by adjudication committee because of data issues (N=1). Therefore, they fell into the Run-in failed group.

**One patient who was entrained with placebo treatment during the Randomized Withdrawal Phase was also previously entrained on placebo during Study 3201.

Source: Post-text Tables 14.1.1 and 14.1.2, Post-text Data Listings 16.2.1, 16.2.6.1, 16.2.6.3, and 16.2.6.6

Source: Figure 2 of sponsor's Clinical study Report

Datasets Analyzed

Data sets analyzed for the Run-in Phase and the Randomized Withdrawal Phase are presented in Table 6. In the Run-in Phase, all 57 patients received at least one dose of study drug and were included in the Safety Population. In the Randomized Withdrawal Phase, 20 patients received at

least one dose of study drug and were included in the Safety Population. The ITT Population included 20 randomized patients who had a τ value calculated post-randomization.

Table 6: Analysis Population Summary -- All Enrolled Patients

	Not Randomized (N=37) n (%)	Placebo (N=10) n (%)	Tasimelteon 20 mg (N=10) n (%)	Total (N=57) n (%)
Enrolled	**37 (100.0)**	**10 (100.0)**	**10 (100.0)**	**57 (100.0)**
Did not receive run-in drug	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Safety population for run-in phase	37 (100.0)	10 (100.0)	10 (100.0)	57 (100.0)
Randomized	**0 (0.0)**	**10 (100.0)**	**10 (100.0)**	**20 (35.1)**
Did not receive study drug	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Safety population for the randomized withdrawal phase	0 (0.0)	10 (100.0)	10 (100.0)	20 (35.1)
Intent-to-treat population	**0 (0.0)**	**10 (100.0)**	**10 (100.0)**	**20 (35.1)**

Percentages were based on the total number of patients enrolled (N).
Source: Post-text Table 14.1.3

Source: Table 11 of sponsor's Clinical Study Report

Demographic and Other Baseline Characteristics

Table 7 summarizes demographics and baseline characteristics for all enrolled patients in the Run-in Phase and the ITT Population in the Randomized Withdrawal Phase. It seems that there were no major differences between the placebo and tasimelteon treatment groups in regard to demographic and baseline characteristics in the Randomized Withdrawal Phase.

Table 7: Demographic and Baseline Characteristics – All Enrolled Patients and ITT population

	Not Randomized[1] (N=37)	Randomized Withdrawal Phase[2]			Total (N=57)
		Placebo (N=10) n (%)	Tasimelteon (N=10) n (%)	Total (N=20) n (%)	
Age at screening (years)					
n	37	10	10	20	57
Mean (SD)	52.4 (13.18)	52.1 (12.01)	51.3 (12.87)	51.7 (12.12)	52.1 (12.71)
Age category (n, %)					
18 to 40 years	9 (24.3)	2 (20.0)	1 (10.0)	3 (15.0)	12 (21.1)
41 to 65 years	24 (64.9)	8 (80.0)	7 (70.0)	15 (75.0)	39 (68.4)
>65 years	4 (10.8)	0 (0.0)	2 (20.0)	2 (10.0)	6 (10.5)
Gender (n, %)					
Male	21 (56.8)	6 (60.0)	6 (60.0)	12 (60.0)	33 (57.9)
Female	16 (43.2)	4 (40.0)	4 (40.0)	8 (40.0)	24 (42.1)
Race (n, %)					
American Indian or Alaska Native	1 (2.7)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.8)
Asian	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Black or African or African American	5 (13.5)	1 (10.0)	0 (0.0)	1 (5.0)	6 (10.5)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
White	31 (83.8)	9 (90.0)	9 (90.0)	18 (90.0)	49 (86.0)
Other	0 (0.0)	0 (0.0)	1 (10.0)	1 (5.0)	1 (1.8)

	Not Randomized[1] (N=37)	Randomized Withdrawal Phase[2]			Total (N=57)
		Placebo (N=10) n (%)	Tasimelteon (N=10) n (%)	Total (N=20) n (%)	
Ethnicity (n, %)					
Hispanic or Latino	0 (0.0)	1 (10.0)	1 (10.0)	2 (10.0)	2 (3.5)
Not Hispanic or Latino	37 (100.0)	9 (90.0)	9 (90.0)	18 (90.0)	55 (96.5)
Weight at screening (kg)					
n	37	10	10	20	57
Mean (SD)	81.16 (14.210)	80.70 (16.028)	86.21 (13.362)	83.46 (14.638)	81.96 (14.273)
Body mass index at screening (kg/m^2)					
n	37	10	10	20	57
Mean (SD)	29.35 (4.025)	28.04 (3.559)	29.24 (3.252)	28.64 (3.375)	29.10 (3.794)
Average of LQ-nTST during run-in phase (hours)					
n	31	10	10	20	51
Mean (SD)	4.259 (1.3773)	5.344 (1.3608)	5.512 (0.9390)	5.428 (1.1412)	4.717 (1.4018)
Median	4.282	5.267	5.356	5.327	4.999
Average of UQ-dTSD during run-in phase (hours)					
n	30	10	10	20	50
Mean (SD)	1.500 (1.2761)	0.994 (0.9025)	0.956 (1.1055)	0.975 (0.9824)	1.290 (1.1855)
Median	1.153	0.891	0.869	0.891	0.990
Average of nTST during run-in phase (hours)					
n	31	10	10	20	51
Mean (SD)	5.899 (0.9224)	6.801 (1.0267)	6.812 (0.8605)	6.806 (0.9220)	6.255 (1.0167)
Median	6.059	6.946	6.919	6.933	6.290
Average of dTSD during run-in phase (hours)					
n	30	10	10	20	50
Mean (SD)	0.607 (0.6045)	0.353 (0.2809)	0.340 (0.4479)	0.346 (0.3639)	0.503 (0.5331)
Median	0.355	0.310	0.268	0.293	0.323

	Not Randomized[1] (N=37)	Randomized Withdrawal Phase[2]			Total (N=57)
		Placebo (N=10) n (%)	Tasimelteon (N=10) n (%)	Total (N=20) n (%)	
MoST during run-in phase (hours)					
n	31	10	10	20	51
Mean (SD)	3.157 (0.8984)	2.906 (1.0147)	3.631 (0.9631)	3.269 (1.0323)	3.201 (0.9446)
Median	3.188	3.080	3.787	3.600	3.214
τ (aMT6s)					
n	28	10	10	20	48
Mean (SD)	24.52 (0.329)	23.99 (0.033)	24.00 (0.043)	24.00 (0.038)	24.30 (0.362)
Median	24.49	23.99	24.00	23.99	24.19
τ (cortisol)					
n	20	10	10	20	40
Mean (SD)	24.48 (0.267)	23.99 (0.080)	24.01 (0.085)	24.00 (0.081)	24.24 (0.311)
Median	24.44	24.00	24.01	24.00	24.12
Previous tasimelteon exposure before run-in phase (n, %)					
Yes	18 (48.6)	7 (70.0)	9 (90.0)	16 (80.0)	34 (59.6)
No	19 (51.4)	3 (30.0)	1 (10.0)	4 (20.0)	23 (40.4)

[1] Demographics and baseline characteristics for the Run-in Phase (All Enrolled Patients).
[2] Demographics and baseline characteristics for the Randomized Withdrawal Phase (ITT Population).
aMT6s = 6-sulfatoxymelatonin; dTSD = daytime total sleep duration; ITT = intent-to-treat; LQ-nTST = lower quartile of subjective nighttime total sleep time; MoST = midpoint of sleep timing; nTST = nighttime total sleep time; UQ-dTSD = upper quartile of subjective daytime total sleep duration; SD = standard deviation.
Source: Post-text Tables 14.1.4 and 14.1.5

Source: Table 13 of sponsor's Clinical Study Report

3.2.2.6 Sponsor's Primary Efficacy Results

The ITT population was utilized for all primary and secondary efficacy endpoints. Table 8 presents a summary of the results of the primary efficacy endpoint. The primary efficacy endpoint was the proportion of non-entrainment of the circadian melatonin rhythm as measured by urinary aMT6s. Non-entrainment was defined as having a post-baseline τ value ≥24.1 or the lower bound of the 95% CI >24.0. The proportion of Non-24 patients who became non-entrained to a 24-hour day after randomization to tasimelteon was statistically significantly less than the proportion of Non-24 patients who became non-entrained after randomization to placebo treatment (% difference = -70.0; p = 0.0026).

Table 8: Summary of Primary Efficacy Endpoint (ITT Population)

Category Statistic	Placebo n/N′ (%)	Tasimelteon 20 mg n/N′ (%)	P-value[1]	P-value[2]
Primary analysis				
Non-entrainment rate (aMT6s)[3]	8/10 (80.0)	1/10 (10.0)	0.0026	0.0055
Sensitivity analyses				
Non-entrainment rate (aMT6s) [3,4]	8/9 (88.9)	1/10 (10.0)	0.0007	0.0011

[1] P-value was based on Barnard's Exact Test, two-sided.
[2] P-value was based on Fisher's Exact Test, two-sided.
[3] Not entrained was defined as having a post-baseline τ value ≥24.1 or the lower bound of the 95% CI >24.0.
[4] This analysis excluded one patient who was randomized to the placebo arm of Study 3201 and was subsequently categorized as entrained.
aMT6s = 6-sulfatoxymelatonin; ITT = intent-to-treat; N′ = number of patients with measurements of τ at specified time point; % = 100*number/N′ of post-baseline time point; Study 3201 = Study VP-VEC-162-3201.
Source: Post-text Tables 14.2.1 and 14.2.2

Source: Table 14 of sponsor's Clinical Study Report

3.2.2.7 Sponsor's Secondary Efficacy Results

Table 9 presents a summary of the results of the secondary efficacy endpoints. Tasimelteon was superior to placebo in the proportion of patients who were non-entrained (as measured by cortisol), the time (circadian and actual) to first relapse event, and sleep/wake parameters (nTST, LQ-nTST, dTSD, UQ-dTSD, and MoST). Please note that all p-values in this table are nominal p-values.

APPEARS THIS WAY ON ORIGINAL

Table 9: Summary of Secondary Efficacy Endpoints (ITT Population)

Category Statistic	Placebo (N=10)	Tasimelteon 20 mg (N=10)	Treatment Difference (Tasimelteon minus Placebo)	P-value
Circadian time to first relapse[1] (median days)[2]	24.7	NE	NA	0.0907
Actual time to first relapse[1] (median weeks)[2]	4.0	NE	NA	0.1481
Circadian time to first relapse (alternate definition)[3] (median)[2]	32.4	NE	NA	0.0181
Non-entrainment rate (cortisol)[4] n (%)[5]	8 (80.0)	2 (20.0)	-60.0	0.0118
Non-entrained and ≥30 minute decrement in nTST[4] n (%)[5]	5 (50.0)	1 (10.0)	-40.0	0.0623
nTST, LS mean hours (LS mean minutes)[6]	-0.74 (-44.49)	-0.20 (-12.23)	0.54 (32.26)	0.1315
LQ-nTST, LS mean hours (LS mean minutes)[6]	-1.23 (-73.74)	-0.11 (-6.74)	1.12 (67.00)	0.0233
dTSD, LS mean hours (LS mean minutes)[6]	0.30 (17.85)	-0.05 (-3.12)	-0.35 (-20.97)	0.0547
UQ-dTSD, LS mean hours (LS mean minutes)[6]	0.83 (49.95)	-0.16 (-9.31)	-0.99 (59.25)	0.0266
MoST, LS mean hours (LS mean minutes)[6]	-0.27 (-16.05)	0.33 (19.99)	0.60 (36.04)	0.0108

[1] Time to relapse with relapse defined as a 45 minute or greater decrement in the weekly average of nTST compared to the Run-in Phase.
[2] Log-rank p-value was based on Kaplan-Meier.
[3] Time to relapse with relapse defined as a 60 minute or greater decrement in the weekly average of nTST compared to the Run-in Phase.
[4] Not entrained was defined as having a post-baseline τ value ≥24.1 or the lower bound of the 95% CI >24.0.
[5] P-value was based on Barnard's Exact Test, two-sided.
[6] P-value was based on analysis of covariance model.
CI = confidence interval; dTSD = daytime total sleep duration; ITT = intent-to-treat; LQ-nTST = lower quartile of subjective nighttime total sleep time; LS mean = least-squares mean; MoST = midpoint of sleep timing; NA = not applicable; NE = non-estimable (there were no observed relapses during the 8 week study); nTST = nighttime total sleep time; UQ-dTSD = upper quartile of subjective daytime total sleep duration.
Source: Post-text Tables 14.2.3, 14.2.4.1, 14.2.5.1, 14.2.6.1, 14.2.7, 14.2.8, 14.2.9, 14.2.10, 14.2.11, 14.2.12, 14.2.13, 14.2.14, 14.2.15, 14.2.16, 14.2.17

Source: Table 15 of sponsor's Clinical Study Report

3.2.3 REVIEWER'S ANALYSIS

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The analyses presented in this section were conducted by this reviewer and they are post-hoc exploratory analyses without multiplicity adjustment.

3.2.3.1 Summary of Important Events Related to Statistical Analysis in Study 3201

Table 10 presents the summary of important events that occurred in Study 3201. In the original protocol, the primary endpoint proposed by the sponsor was nTST and the proposed sample size was 160 patients, based on the postulated mean treatment difference of 39 minutes and standard deviation of 66 minutes. In Amendment 6 submitted to the Agency, the sample size was changed

from 160 to 100 patients, based on the new postulated mean treatment difference of 30 minutes and standard deviation of 45 minutes. In Amendment 9, the primary endpoint was changed to entrainment and the sample size was reduced to 84 patients. At the time of Amendment 9 (May 21, 2012), 95% of the patients were randomized and 56% of the patients completed the study. Amendment 11 was dated on December 11, 2012 and the trial data was unblinded on December 12, 2012. It is unclear to the Agency how much these changes might have impacted the trial results.

Table 10: Summary of Important Events Related to Statistical Analysis in Study 3201

Study 3201 Important Events	Date	Primary Endpoint	Sample Size	Note
Original	5/24/2010	nTST	160	postulated mean difference =39 mins and std=66 mins
First Patient Enrolled	8/25/2010			
Amendment 6	8/8/2011	nTST	100	postulated mean difference =30 mins and std=45 mins
Amendment 9	5/21/2012	Entrainment	84	80/84 patients (95%) randomized; 47/84 patients (56%) completed
Last Patient Completed	10/29/2012			
Amendment 11	12/11/2012	Entrainment	84	
Data Unblinding	12/12/2012	Entrainment	84	

Source: Reviewer's Analysis

3.2.3.2 Summary of Statistical Issues

Study 3201

There are three main statistical issues for this study:

Primary Efficacy endpoint(s)

Sponsor's primary efficacy analyses were based on a melatonin-based biomarker endpoint "entrainment". However, throughout the development program for tasimelteon, the Agency repeatedly disagrees that this biomarker is adequate to assess the efficacy for this indication and requires clinically meaningful primary endpoint(s). No agreement was reached between the Agency and the sponsor regarding the primary endpoint(s). The Agency has decided that the efficacy evaluation for tasimelteon should be based on the following clinical endpoints: LQ-nTST, UQ-dTSD, MoST, CGIC, nTST and dTSD.

Analysis Populations

For Study 3201, 84 patients were randomized (n=84). The ITT Population defined by the sponsor included all patients randomized into the study that had τ calculated post-randomization (thereafter referred as Sponsor ITT, n=78). The Analysis Population defined by the sponsor included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization (thereafter referred as Sponsor Analysis Population, n=72). The sponsor's ITT and Analysis Population were selected after the randomization and they are non-randomized subsets. Six patients in the Randomized Population without τ calculated post-randomization were excluded from the Sponsor ITT and six patients in the Sponsor ITT were excluded from the Sponsor Analysis Population due to less than 70% of 1 circadian cycle of nTST data reported during screening and/or post-randomization. Thus, 12 patients in total were excluded from the 84 patients randomized. Please refer to Table 2 for Analysis Population Summary. In sponsor's efficacy analyses, the Sponsor ITT population was utilized for all circadian rhythm related outcomes and the Analysis Population was used for analyses of all other endpoints including the step-down primary and all other efficacy analyses. However, the 84 randomized patients all took study medication and had at least one baseline and postbaseline assessment. Based on the intent-to-treat principal, all the 84 patients should be included in the ITT population (n=84). This ITT population is a randomized population. The efficacy analyses conducted by this reviewer are based on ITT population. In Section 3.2.3.4, this reviewer will present information regarding the 12 patients excluded from the efficacy analyses by the sponsor and explain why these 12 patients should be included in the efficacy analyses.

Analysis Methods

In sponsor's efficacy analyses for the clinical endpoints, ANCOVA model was used which includes baseline value as a covariate and treatment group and pooled sites as factors. However, due to small sample size, non-normal distribution of the data, and some heterogeneity in variances, this reviewer thinks permutation ANCOVA is more appropriate than ANCOVA. In addition, in sponsor's ANCOVA analysis for Study 3201, study site was included as a factor, but the Study Report shows that the randomization was not stratified by study site. Normally, if the randomization isn't stratified by study site, in order to comply with the trial design, the site is not necessarily included in the analysis model. Therefore, this reviewer thinks permutation ANCOVA without site is more appropriate than ANCOVA with site.

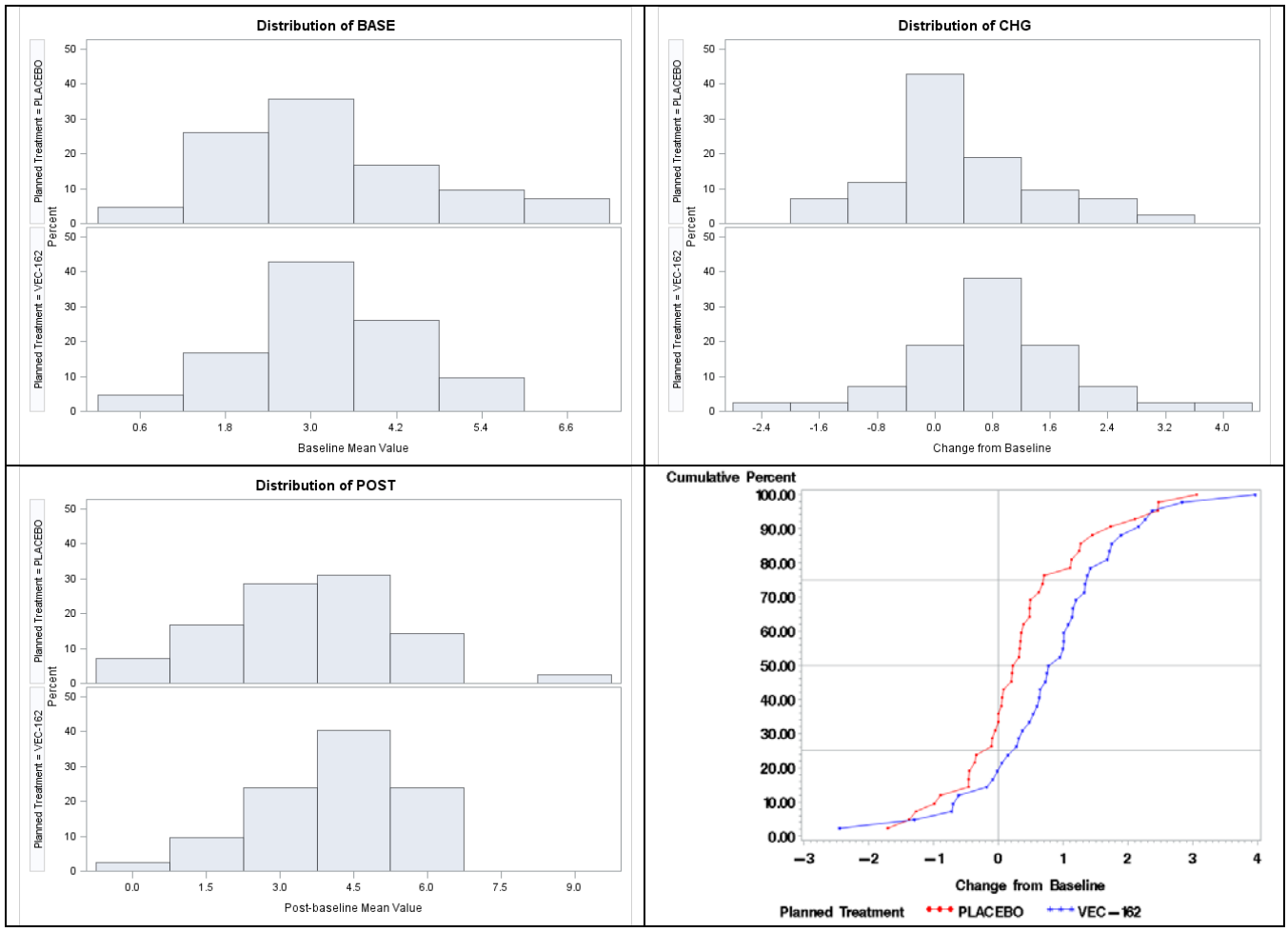
Study 3203

For this study, 20 randomized patients all received the study medication and completed the study, thus all 20 patients were included in the ITT population. The issues related to primary endpoint and analysis method are similar to Study 3201, but please note that for Study 3203, the randomization was not stratified by site and site was not a factor in sponsor's ANCOVA analysis.

3.2.3.3 Graphic Presentation of Clinical Endpoints

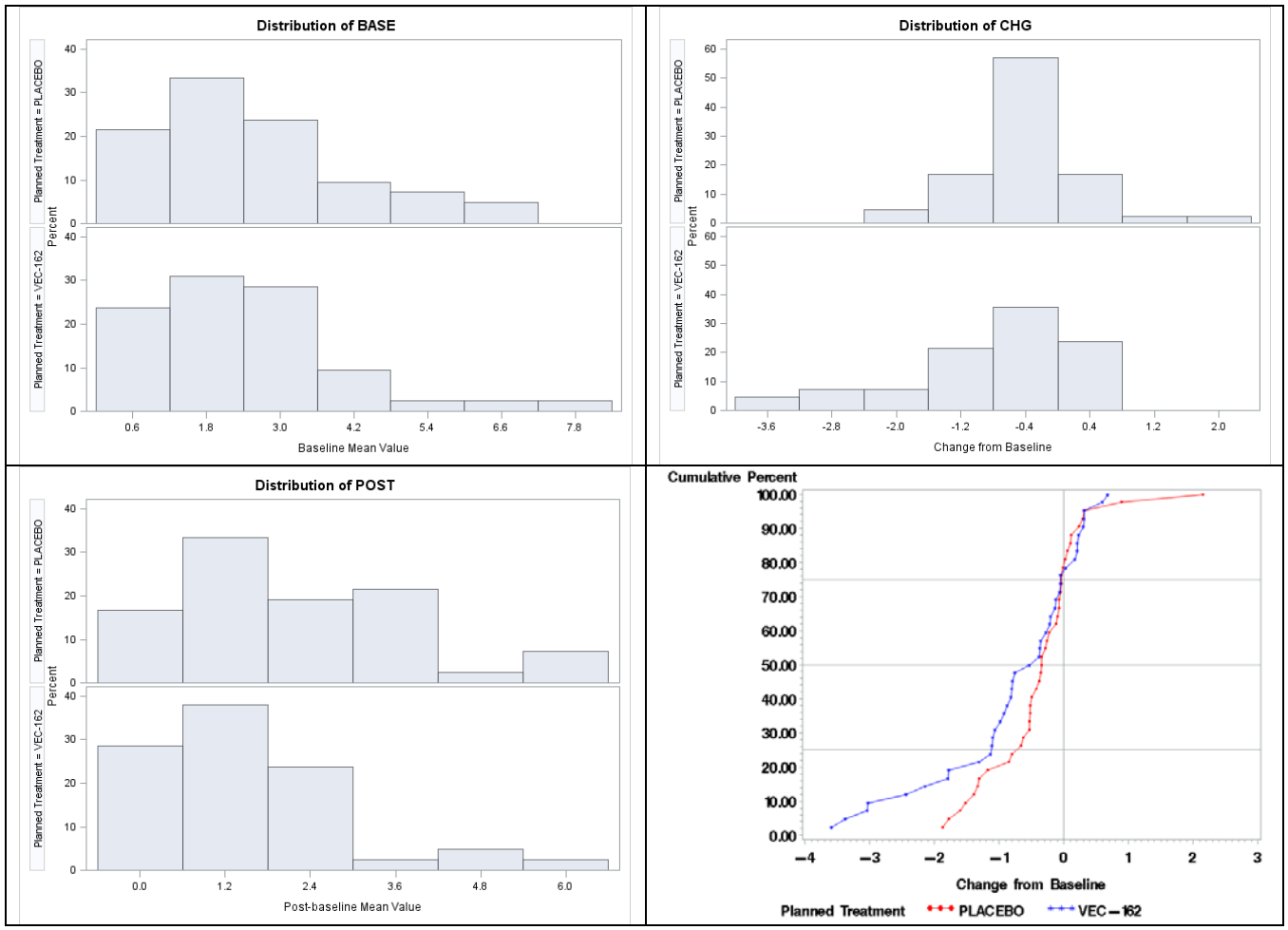
In order to visually examine the distribution of the clinical endpoints, this reviewer generated histograms for each of the clinical endpoints in both studies. The following figures present the Cumulative Distribution Functions (CDFs) and the histograms for baseline, postbaseline and changes from baseline for each of the clinical endpoints in each study.

Figure 6: Histograms and CDF for LQ-nTST (Study 3201, FDA ITT, n=84)



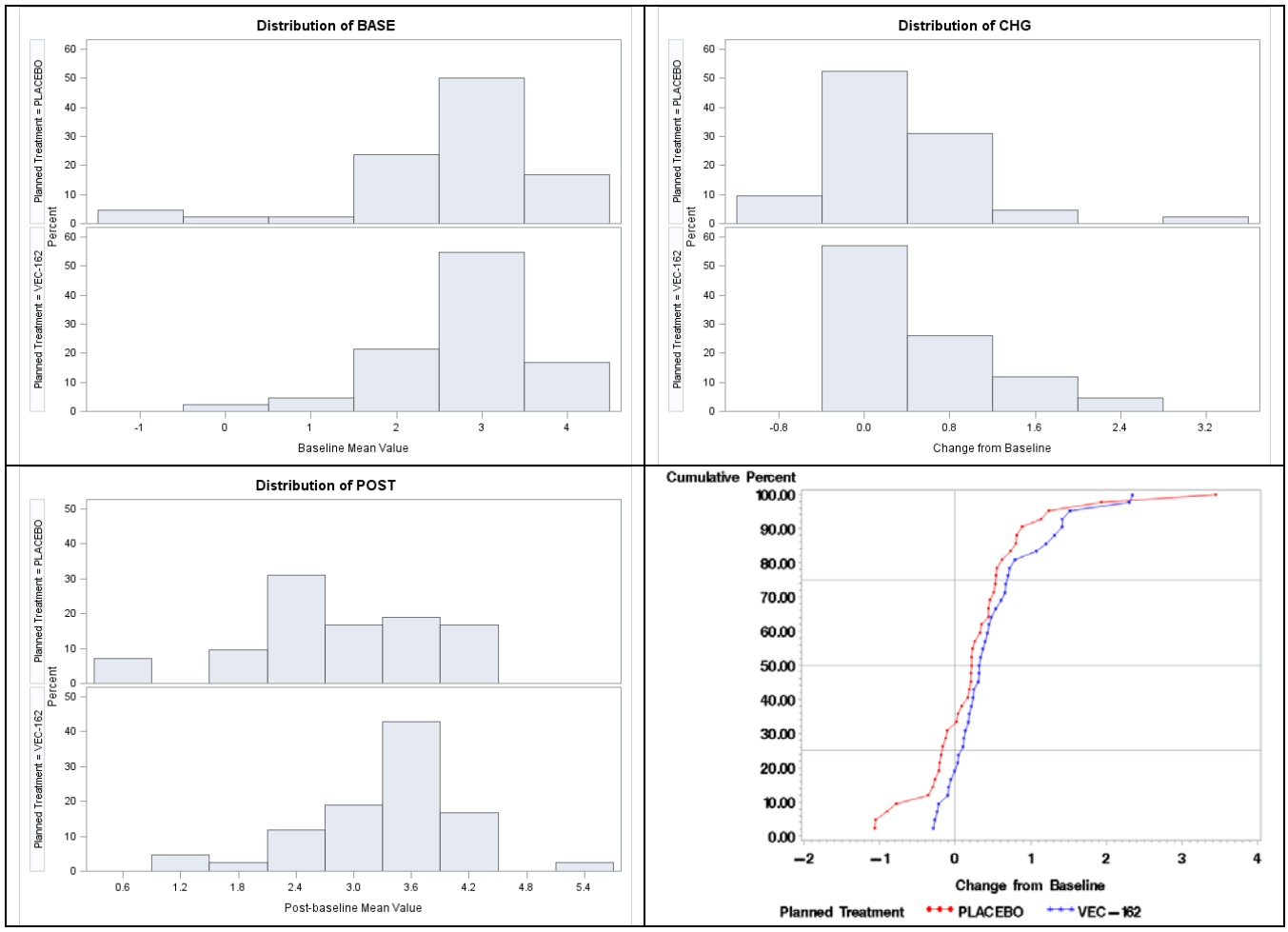
Source: Reviewer's Analysis
*: For LQ-nTST, larger values indicate longer nighttime total sleep time and positive change from baseline shows improvement in nighttime total sleep time.

Figure 7: Histograms and CDF for UQ-dTSD (Study 3201, FDA ITT, n=84)



Source: Reviewer's Analysis
*: For UQ-dTSD, smaller values indicate shorter daytime total sleep duration and negative change from baseline shows improvement in daytime total sleep duration.

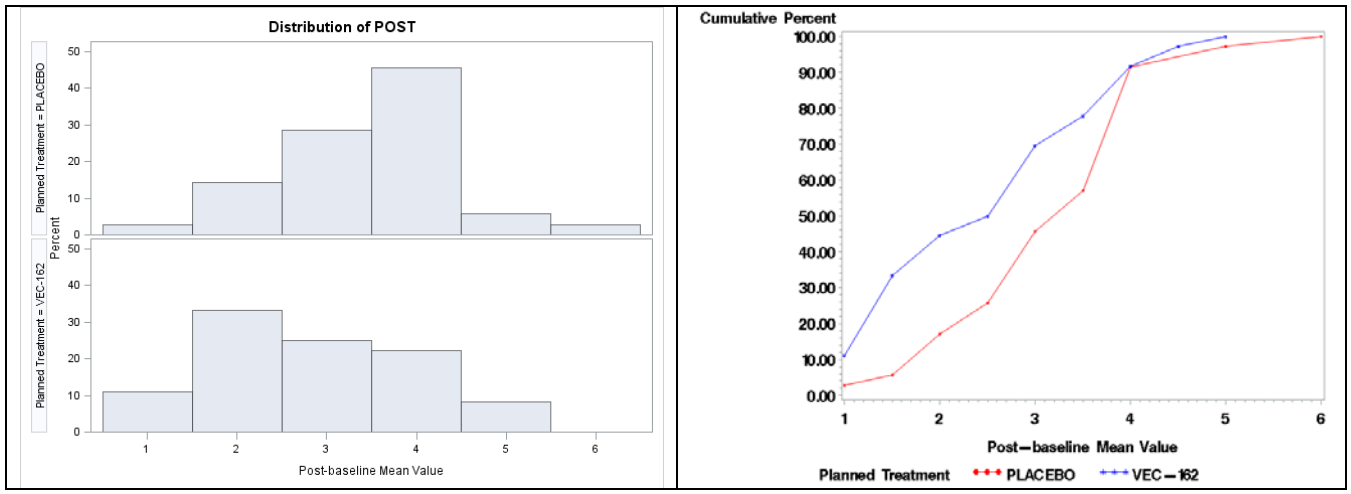
Figure 8: Histograms and CDF for MoST (Study 3201, FDA ITT, n=84)



Source: Reviewer's Analysis

*: For MoST, an individual who has late afternoon naps and early morning awakenings would have a small MoST value or a negative number, and a positive change from baseline indicates improvement in MoST.

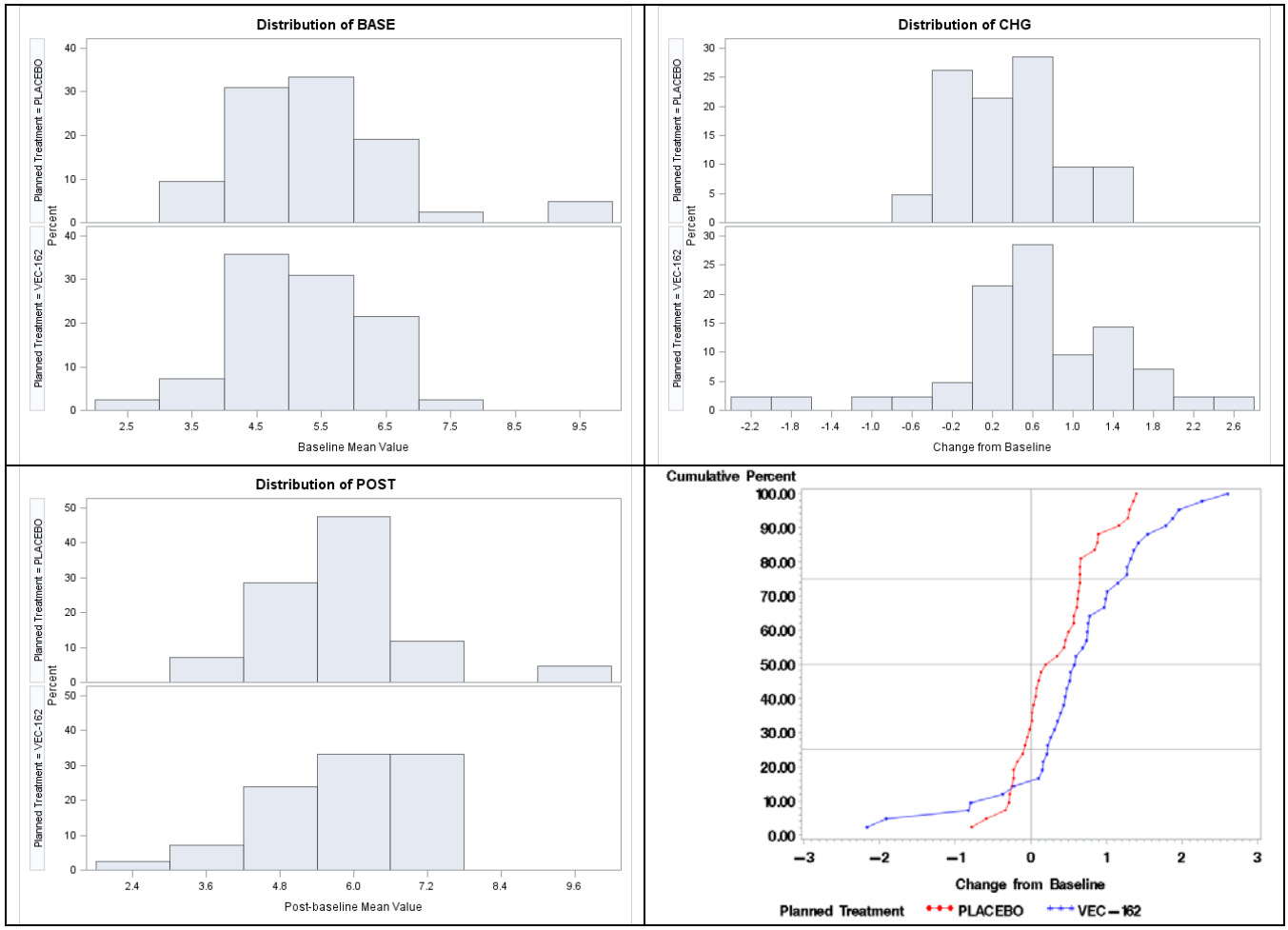
Figure 9: Histograms and CDF for CGIC (Study 3201, FDA ITT, n=84)



Source: Reviewer's Analysis
*: CGI-C is a 7-point rating scale: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 = minimally worse; 6 = much worse; or 7 = very much worse.

APPEARS THIS WAY ON ORIGINAL

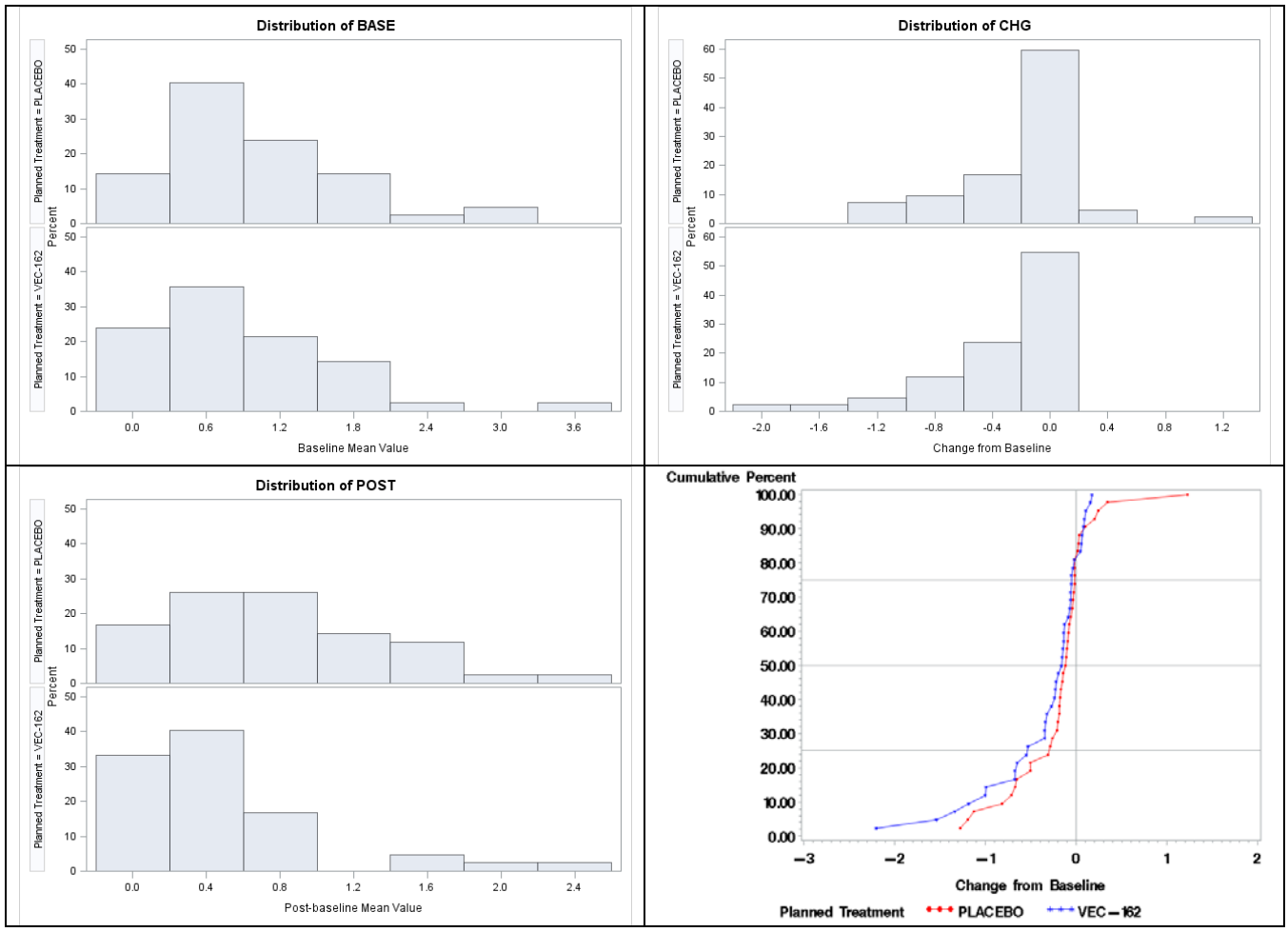
Figure 10: Histograms and CDF for nTST (Study 3201, FDA ITT, n=84)



Source: Reviewer's Analysis

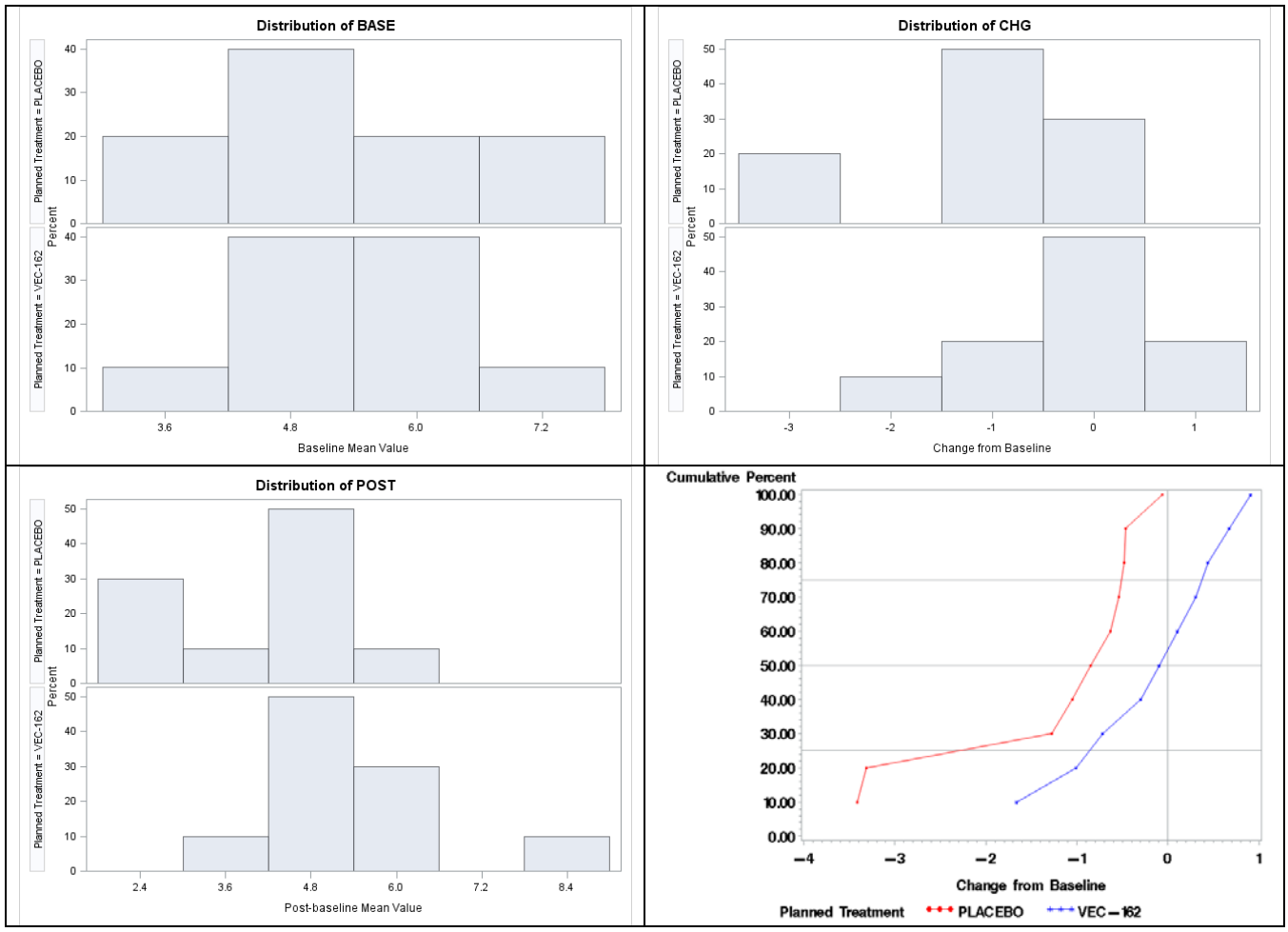
*: For nTST, larger values indicate longer nighttime total sleep time and positive change from baseline shows improvement in nighttime total sleep time.

Figure 11: Histograms and CDF for dTSD (Study 3201, FDA ITT, n=84)



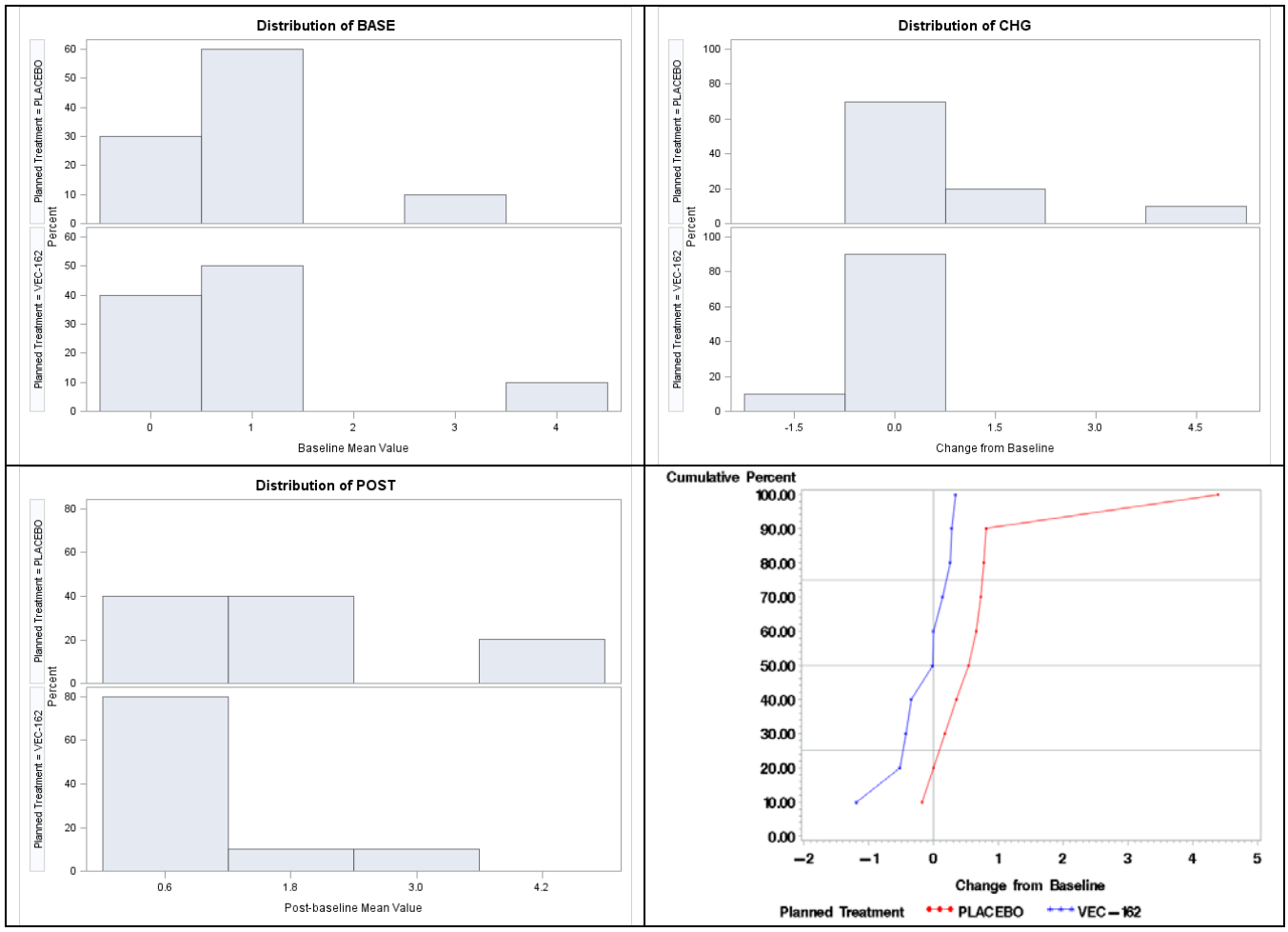
Source: Reviewer's Analysis
*: For dTSD, smaller values indicate shorter daytime total sleep duration and negative change from baseline shows improvement in daytime total sleep duration.

Figure 12: Histograms and CDF for LQ-nTST (Study 3203, n=20)



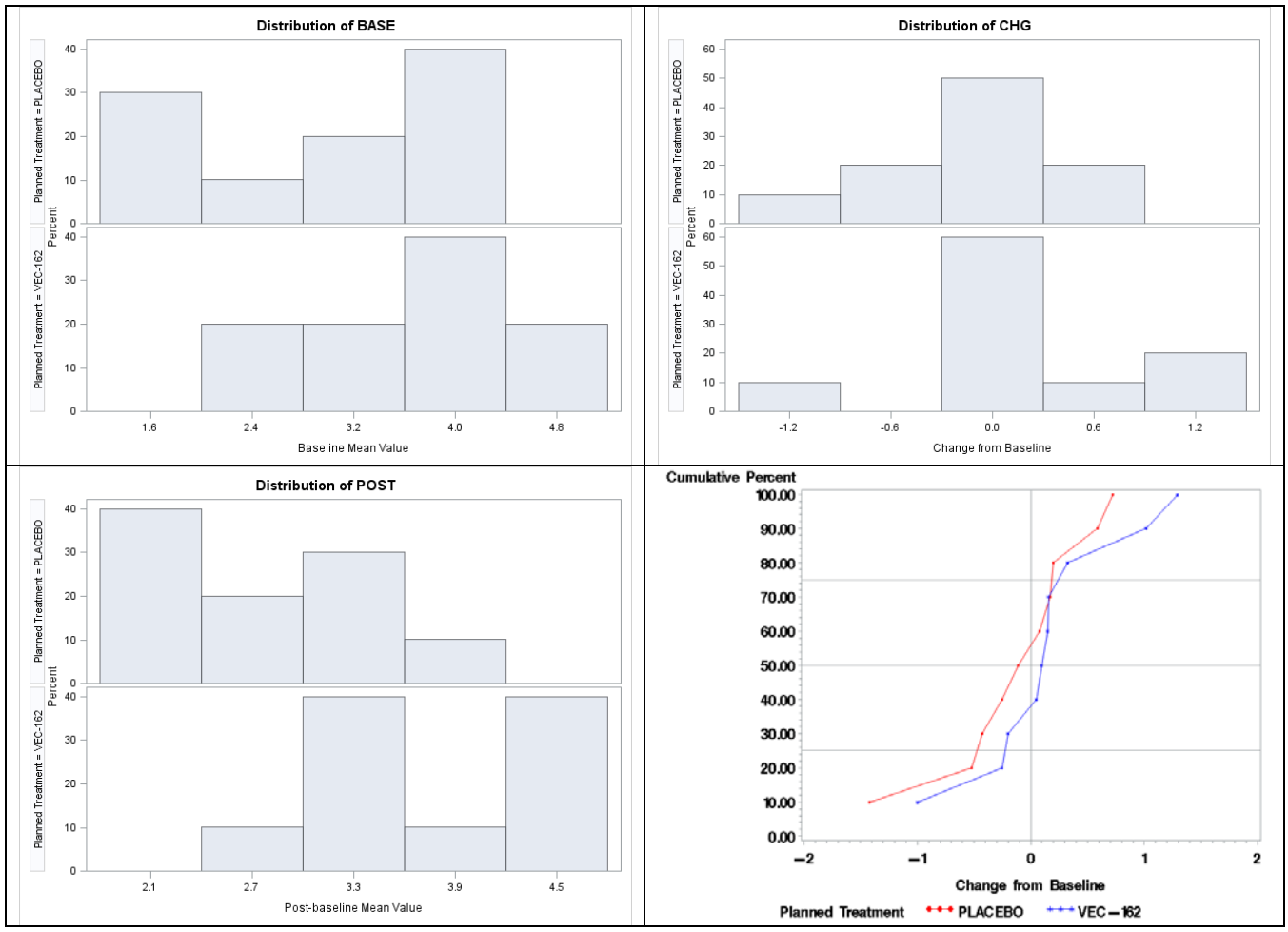
Source: Reviewer's Analysis
*: For LQ-nTST, larger values indicate longer nighttime total sleep time and positive change from baseline shows improvement in nighttime total sleep time.

Figure 13: Histograms and CDF for UQ-dTSD (Study 3203, n=20)



Source: Reviewer's Analysis
*: For UQ-dTSD, smaller values indicate shorter daytime total sleep duration and negative change from baseline shows improvement in daytime total sleep duration.

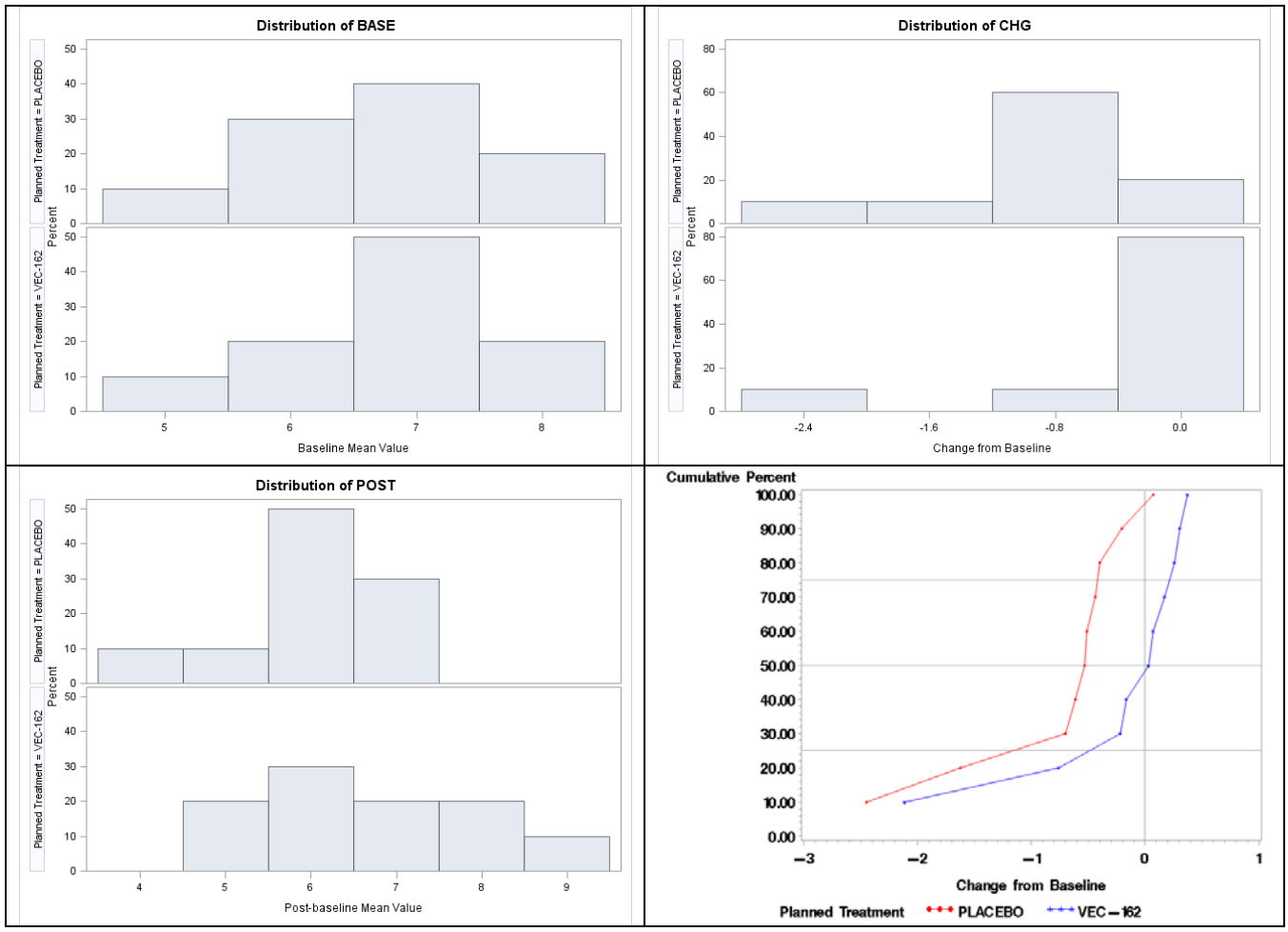
Figure 14: Histograms and CDF for MoST (Study 3203, n=20)



Source: Reviewer's Analysis

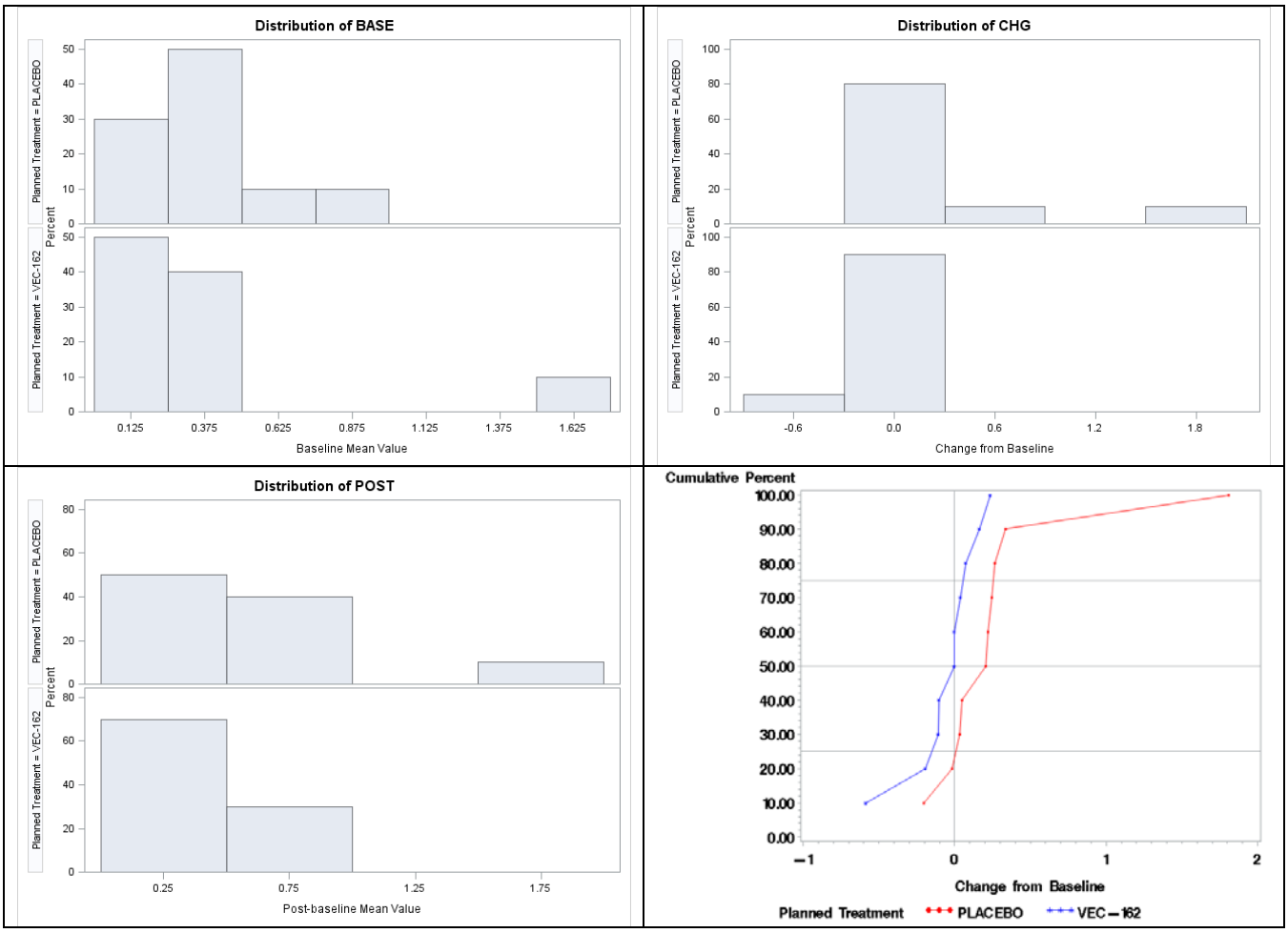
*: For MoST, an individual who has late afternoon naps and early morning awakenings would have a small MoST value or a negative number, and a positive change from baseline indicates improvement in MoST.

Figure 15: Histograms and CDF for nTST (Study 3203, n=20)



Source: Reviewer's Analysis
*: For nTST, larger values indicate longer nighttime total sleep time and positive change from baseline shows improvement in nighttime total sleep time.

Figure 16: Histograms and CDF for dTSD (Study 3203, n=20)



Source: Reviewer's Analysis
*: For dTSD, smaller values indicate shorter daytime total sleep duration and negative change from baseline shows improvement in daytime total sleep duration.

3.2.3.4 The Twelve Patients Excluded from Efficacy Analysis by Sponsor for Study 3201

Sponsor's efficacy analyses for clinical endpoints were based on Analysis Population. The Analysis Population included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization. Based on this definition, 12 patients in the Randomized Population were excluded from efficacy analysis for clinical endpoints. Table 11 presents the cycle length, % of one circadian cycle of nTST at baseline and postbaseline for the 12 patients. Among these 12 patients, 10 patients had more than 50% of 1 circadian cycle of nTST at baseline and 6 patients had more than 50% of 1 circadian cycle of nTST at postbaseline. It seems that the data are not extremely sparse for these 12 patients and this reviewer thinks they should be included in the efficacy analysis.

Table 11: The 12 Patients Excluded from Efficacy Analysis (Study 3201)

Subject ID	Treatment Group	Cycle Length	% Cycle Base	% Cycle Postbaseline
(b) (6)	PLACEBO	32	275.0	53.1
	PLACEBO	54	148.1	13.0
	PLACEBO	54	94.4	11.1
	PLACEBO	58	36.2	129.3
	PLACEBO	59	113.6	27.1
	PLACEBO	79	65.8	160.8
	PLACEBO	81	60.5	145.7
	PLACEBO	96	49.0	79.2
	VEC-162	72	70.8	2.8
	VEC-162	85	116.5	23.5
	VEC-162	94	66.0	107.4
	VEC-162	97	84.5	36.1

Source: Reviewer's Analysis

The table below (Table 12) provides the descriptive statistics for the 12 patients excluded from the Sponsor Analysis Population versus the 72 patients in the Sponsor Analysis Population. It appears that the inclusion of the 12 patients neither consistently strengthen nor consistently weaken the treatment effect for different clinical endpoints.

Table 12: Descriptive Statistics for the 12 Patients Excluded from Sponsor Analysis Population

Endpoint	Treatment	Flag*	N	Mean	Median	Std Dev	Std Error	Min.	Max.
1. LQ-nTST	PLACEBO	N	8	0.56	0.42	1.92	0.68	-1.71	3.06
		Y	34	0.32	0.22	0.71	0.12	-0.99	2.11
	VEC-162	N	4	-0.03	0.46	1.69	0.85	-2.44	1.42
		Y	38	0.92	0.98	1.04	0.17	-1.29	3.96
2. UQ-dTSD	PLACEBO	N	8	-0.33	-0.53	1.20	0.43	-1.78	2.15
		Y	34	-0.41	-0.31	0.60	0.10	-1.87	0.90
	VEC-162	N	4	-1.54	-1.58	2.25	1.13	-3.58	0.60
		Y	38	-0.69	-0.45	0.91	0.15	-3.03	0.68
3. MoST	PLACEBO	N	8	0.13	-0.52	1.56	0.55	-1.06	3.44
		Y	34	0.32	0.22	0.47	0.08	-0.36	1.94
	VEC-162	N	4	0.37	0.45	0.33	0.16	-0.09	0.67
		Y	38	0.51	0.32	0.65	0.11	-0.29	2.35
4. CGIC	PLACEBO	N	2	2.25	2.25	0.35	0.25	2.00	2.50
	VEC-162	Y	33	3.41	3.50	1.03	0.18	1.00	6.00
		Y	36	2.63	2.75	1.17	0.19	1.00	5.00
5. nTST	PLACEBO	N	8	-0.00	-0.10	0.58	0.21	-0.78	0.89
		Y	34	0.40	0.39	0.51	0.09	-0.34	1.40
	VEC-162	N	4	-0.21	0.15	1.40	0.70	-2.16	1.02
		Y	38	0.71	0.65	0.87	0.14	-1.91	2.61
6. dTSD	PLACEBO	N	8	-0.05	-0.19	0.55	0.20	-0.65	1.22
		Y	34	-0.24	-0.11	0.39	0.07	-1.28	0.34
	VEC-162	N	4	-0.65	-0.61	0.91	0.45	-1.54	0.16
		Y	38	-0.31	-0.16	0.45	0.07	-2.20	0.18

Source: Reviewer's Analysis
*Flag: N=the patient isn't in the Sponsor Analysis Population; Y=the patient is in the Sponsor Analysis Population.

3.2.3.5 Results of ANCOVA Analysis for Clinical Endpoints on ITT Population

The three tables below (Table 13, Table 14 and Table 15) present the results of ANCOVA analysis on ITT population for clinical endpoints. In sponsor's ANCOVA analysis for Study

3201, baseline was included as a covariate if applicable and the pooled sites as a factor (variable name: SITEGR3). However, since the randomization wasn't stratified by site, this reviewer thinks it isn't necessary to include site as a factor in the analysis model. Based on the results of ANCOVA analysis without site, it seems that for both studies the nominal p-values were statistically significant or marginally significant for LQ-nTST, UQ-dTSD, MoST, CGIC and dTSD. The nominal p-value for nTST wasn't statistically significant in either of the studies.

Table 13: LS Mean in Hours and P-value from ANCOVA with Site (Study 3201, ITT, n=84)

Endpoints	Placebo	Tasimelteon	P-value
LQ-nTST	0.29	0.89	0.0232
UQ-dTSD	-0.28	-0.84	0.0031
MoST	0.22	0.56	0.0229
CGIC[1]	3.38	2.62	0.0104
nTST	0.32	0.65	0.0658
dTSD	-0.14	-0.39	0.0026

Source: Reviewer's Analysis
[1]:For CGIC, n=71 since CGIC was missing for 13 patients.
*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 14: LS Mean in Hours and P-value from ANCOVA without Site (Study 3201, ITT, n=84)

Endpoints	Placebo	Tasimelteon	P-value
LQ-nTST	0.37	0.83	0.0510
UQ-dTSD	-0.36	-0.81	0.0118
MoST	0.25	0.54	0.0366
CGIC[1]	3.34	2.63	0.0080
nTST	0.35	0.60	0.1149
dTSD	-0.18	-0.36	0.0166

Source: Reviewer's Analysis
[1]:For CGIC, n=71 since CGIC was missing for 13 patients.
*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 15: LS Mean in Hours and P-value from ANCOVA without Site (Study 3203, ITT, n=20)

Endpoints	Placebo	Tasimelteon	P-value
LQ-nTST	-1.23	-0.11	0.0233
UQ-dTSD	0.83	-0.16	0.0266
MoST	0.83	-0.16	0.0266
nTST	-0.74	-0.20	0.1315
dTSD	0.30	-0.05	0.0547

Source: Reviewer's Analysis
*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD and dTSD, smaller values indicate better outcome.

3.2.3.6 Permutation ANCOVA Test for Clinical Endpoints for ITT Population

The following two tables (Table 16 and Table 17) provide the descriptive statistics for clinical endpoints. It seems that there is some heterogeneity in the variances of the clinical endpoints. Furthermore, due to small sample size and non-normal distribution of the clinical endpoints, this reviewer thinks permutation ANCOVA would be more appropriate than ANCOVA.

Table 16: Descriptive Statistics for Clinical Endpoints (Study 3201, ITT, n=84)

Endpoint	Planned Treatment	n	Mean	Std Dev	Variance	Min.	Max.
LQ-nTST	PLACEBO	42	0.37	1.02	1.04	-1.71	3.06
	VEC-162	42	0.83	1.13	1.27	-2.44	3.96
UQ-dTSD	PLACEBO	42	-0.39	0.73	0.54	-1.87	2.15
	VEC-162	42	-0.77	1.08	1.18	-3.58	0.68
MoST	PLACEBO	42	0.28	0.77	0.60	-1.06	3.44
	VEC-162	42	0.50	0.63	0.39	-0.29	2.35
CGIC	PLACEBO	35	3.34	1.04	1.08	1.00	6.00
	VEC-162	36	2.63	1.17	1.36	1.00	5.00
nTST	PLACEBO	42	0.33	0.54	0.30	-0.78	1.40
	VEC-162	42	0.62	0.95	0.90	-2.16	2.61
dTSD	PLACEBO	42	-0.20	0.43	0.18	-1.28	1.22
	VEC-162	42	-0.34	0.50	0.25	-2.20	0.18

Source: Reviewer's Analysis

*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 17: Descriptive Statistics for Clinical Endpoints (Study 3203, ITT, n=20)

Endpoint	Planned Treatment	N	Mean	Std Dev	Variance	Min.	Max.
LQ-nTST	PLACEBO	10	-1.21	1.18	1.40	-3.42	-0.06
	VEC-162	10	-0.13	0.80	0.64	-1.66	0.91
UQ-dTSD	PLACEBO	10	0.82	1.30	1.68	-0.18	4.38
	VEC-162	10	-0.15	0.48	0.23	-1.19	0.34
MoST	PLACEBO	10	-0.10	0.61	0.38	-1.42	0.72
	VEC-162	10	0.16	0.64	0.41	-1.00	1.29
nTST	PLACEBO	10	-0.74	0.74	0.55	-2.45	0.07
	VEC-162	10	-0.20	0.75	0.56	-2.11	0.37
dTSD	PLACEBO	10	0.29	0.56	0.31	-0.20	1.81
	VEC-162	10	-0.05	0.23	0.05	-0.59	0.23

Source: Reviewer's Analysis

*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

The results of permutation ANCOVA are shown in the following three tables (Table 18, Table 19 and Table 20). For each permutation test, the number of repetition is 100,000. The seeds were 3201/3203, 11, 14, 2013 and 100000 since the study number is 3201/3203 and Advisory Committee meeting was scheduled on Nov. 14, 2013.

Table 18: Permutation ANCOVA with Site (Study 3201, ITT, n=84)

Endpoint	ANCOVA with Site	Permutation ANCOVA				
		Seed=3201	Seed=11	Seed=14	Seed=2013	Seed=100000
LQ-nTST	0.0232	0.0236	0.0244	0.0246	0.0242	0.0244
UQ-dTSD	0.0031	0.0032	0.0033	0.0028	0.0030	0.0029
MoST	0.0229	0.0218	0.0219	0.0207	0.0213	0.0219
CGIC	0.0104	0.0138	0.0146	0.0146	0.0138	0.0145
nTST	0.0658	0.0684	0.0663	0.0689	0.0677	0.0677
dTSD	0.0026	0.0024	0.0023	0.0020	0.0025	0.0023

Source: Reviewer's Analysis

Table 19: Permutation ANCOVA without Site (Study 3201, ITT, n=84)

Endpoint	ANCOVA without Site	Permutation ANCOVA				
		Seed=3201	Seed=11	Seed=14	Seed=2013	Seed=100000
LQ-nTST	0.0510	0.0516	0.0519	0.0528	0.0518	0.0527
UQ-dTSD	0.0118	0.0112	0.0117	0.0108	0.0114	0.0112
MoST	0.0366	0.0361	0.0365	0.0363	0.0365	0.0371
CGIC	0.0080	0.0083	0.0083	0.0080	0.0077	0.0082
nTST	0.1149	0.1168	0.1164	0.1163	0.1169	0.1174
dTSD	0.0166	0.0154	0.0162	0.0146	0.0155	0.0157

Source: Reviewer's Analysis

Table 20: Results of Permutation ANCOVA without Site (Study 3203, ITT, n=20)

Endpoint	ANCOVA without Site	Permutation ANCOVA				
		Seed=3203	Seed=11	Seed=14	Seed=2013	Seed=100000
LQ-nTST	0.0233	0.0226	0.0227	0.0234	0.0237	0.0235
UQ-dTSD	0.0266	0.0069	0.0064	0.0074	0.0070	0.0067
MoST	0.0108	0.0063	0.0067	0.0075	0.0068	0.0064
nTST	0.1315	0.1532	0.1530	0.1546	0.1548	0.1523
dTSD	0.0547	0.0224	0.0229	0.0234	0.0224	0.0235

Source: Reviewer's Analysis

It appears that the results of permutation ANCOVA without site is fairly close to those of ANCOVA without site. For LQ-nTST, UQ-dTSD, MoST, CGIC, and dTSD, the nominal p-values are statistically significant or marginally significant for both studies. The p-value for nTST isn't statistically significant for either of the studies.

3.3 Evaluation of Safety

Please read Dr. Jillapalli's review for safety assessment.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age and Geographic Region

4.1.1 STUDY 3201 (SET STUDY)

The following four tables (Table 21, Table 22, Table 23, Table 24) present the results of subgroup analysis by sex, race, age group (<50 and >=50) and country (Germany and US). It seems that the point estimates of treatment effects are all in the right direction for sex, age group and country for all clinical endpoints. For race, the point estimates of treatment effect for LQ-nTST, MoST and

CGIC aren't in the right direction for non-white patients. However, this doesn't raise concerns since there are only 14 non-white patients in this study.

Table 21: Subgroup Analysis by Sex (Study 3201, ITT, n=84)

Endpoint	Sex	Treatment	N	Mean	Median	Std Dev	Std Error
1. LQ-nTST	F	PLACEBO	17	0.22	0.04	0.70	0.17
		VEC-162	18	0.76	0.70	1.19	0.28
	M	PLACEBO	25	0.47	0.34	1.20	0.24
		VEC-162	24	0.89	1.04	1.09	0.22
2. UQ-dTSD	F	PLACEBO	17	-0.57	-0.49	0.52	0.13
		VEC-162	18	-0.74	-0.37	0.99	0.23
	M	PLACEBO	25	-0.27	-0.08	0.84	0.17
		VEC-162	24	-0.79	-0.65	1.17	0.24
3. MoST	F	PLACEBO	17	0.25	0.22	0.68	0.16
		VEC-162	18	0.41	0.28	0.63	0.15
	M	PLACEBO	25	0.31	0.22	0.84	0.17
		VEC-162	24	0.56	0.35	0.63	0.13
4. CGIC	F	PLACEBO	15	3.63	4.00	0.88	0.23
		VEC-162	15	2.60	3.00	0.99	0.25
	M	PLACEBO	20	3.13	3.00	1.12	0.25
		VEC-162	21	2.64	2.00	1.31	0.28
5. nTST	F	PLACEBO	17	0.33	0.35	0.56	0.14
		VEC-162	18	0.45	0.47	0.94	0.22
	M	PLACEBO	25	0.32	0.13	0.54	0.11
		VEC-162	24	0.75	0.75	0.96	0.20
6. dTSD	F	PLACEBO	17	-0.32	-0.19	0.44	0.11
		VEC-162	18	-0.38	-0.16	0.57	0.13
	M	PLACEBO	25	-0.12	-0.08	0.41	0.08
		VEC-162	24	-0.31	-0.17	0.45	0.09

Source: Reviewer's Analysis

*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 22: Subgroup Analysis by Race (Study 3201, ITT, n=84)

Endpoint	Race	Treatment	N	Mean	Median	Std Dev	Std Error
1. LQ-nTST	NON-WHITE	PLACEBO	8	1.52	1.77	1.19	0.42
		VEC-162	6	0.57	0.55	1.37	0.56
	WHITE	PLACEBO	34	0.10	0.21	0.77	0.13
		VEC-162	36	0.88	1.01	1.10	0.18
2. UQ-dTSD	NON-WHITE	PLACEBO	8	-0.64	-0.25	0.79	0.28
		VEC-162	6	-0.65	-0.40	0.95	0.39
	WHITE	PLACEBO	34	-0.33	-0.34	0.72	0.12
		VEC-162	36	-0.79	-0.56	1.12	0.19
3. MoST	NON-WHITE	PLACEBO	8	0.75	0.50	1.23	0.44
		VEC-162	6	0.53	0.52	0.35	0.14
	WHITE	PLACEBO	34	0.17	0.21	0.59	0.10
		VEC-162	36	0.49	0.32	0.66	0.11
4. CGIC	NON-WHITE	PLACEBO	6	2.92	3.00	1.16	0.47
		VEC-162	6	3.17	3.75	1.57	0.64
	WHITE	PLACEBO	29	3.43	3.50	1.02	0.19
		VEC-162	30	2.52	2.50	1.07	0.20
5. nTST	NON-WHITE	PLACEBO	8	0.58	0.54	0.44	0.16
		VEC-162	6	0.73	0.70	1.04	0.42
	WHITE	PLACEBO	34	0.27	0.12	0.55	0.10
		VEC-162	36	0.60	0.59	0.95	0.16
6. dTSD	NON-WHITE	PLACEBO	8	-0.28	-0.17	0.34	0.12
		VEC-162	6	-0.34	-0.15	0.39	0.16
	WHITE	PLACEBO	34	-0.18	-0.12	0.45	0.08
		VEC-162	36	-0.34	-0.18	0.52	0.09

Source: Reviewer's Analysis

*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 23: Subgroup Analysis for Age Group (Study 3201, ITT, n=84)

Endpoint	Age Group	Treatment	N	Mean	Median	Std Dev	Std Error
1. LQ-nTST	<50	PLACEBO	18	0.30	0.14	1.17	0.28
		VEC-162	17	1.19	1.01	1.12	0.27
	>=50	PLACEBO	24	0.42	0.33	0.91	0.19
		VEC-162	25	0.59	0.60	1.08	0.22
2. UQ-dTSD	<50	PLACEBO	18	-0.54	-0.51	0.77	0.18
		VEC-162	17	-1.29	-1.10	1.13	0.28
	>=50	PLACEBO	24	-0.28	-0.31	0.70	0.14
		VEC-162	25	-0.42	-0.20	0.91	0.18
3. MoST	<50	PLACEBO	18	0.38	0.21	1.02	0.24
		VEC-162	17	0.80	0.67	0.79	0.19
	>=50	PLACEBO	24	0.21	0.22	0.53	0.11
		VEC-162	25	0.29	0.25	0.38	0.08
4. CGIC	<50	PLACEBO	15	3.40	3.50	0.87	0.22
		VEC-162	14	2.43	1.75	1.38	0.37
	>=50	PLACEBO	20	3.30	3.25	1.17	0.26
		VEC-162	22	2.75	3.00	1.02	0.22
5. nTST	<50	PLACEBO	18	0.24	0.27	0.53	0.12
		VEC-162	17	0.84	0.76	0.85	0.21
	>=50	PLACEBO	24	0.39	0.27	0.56	0.11
		VEC-162	25	0.47	0.48	1.00	0.20
6. dTSD	<50	PLACEBO	18	-0.33	-0.19	0.47	0.11
		VEC-162	17	-0.53	-0.34	0.59	0.14
	>=50	PLACEBO	24	-0.10	-0.11	0.38	0.08
		VEC-162	25	-0.21	-0.09	0.40	0.08

Source: Reviewer's Analysis

*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

Table 24: Subgroup Analysis for Country (Study 3201, ITT, n=84)

Endpoint	Country	Treatment	N	Mean	Median	Std Dev	Std Error
1. LQ-nTST	Germany	PLACEBO	4	-0.99	-1.14	0.74	0.37
		VEC-162	4	1.27	1.44	0.70	0.35
	US	PLACEBO	38	0.51	0.33	0.94	0.15
		VEC-162	38	0.79	0.76	1.16	0.19
2. UQ-dTSD	Germany	PLACEBO	4	0.20	-0.26	1.35	0.68
		VEC-162	4	-0.53	-0.49	0.70	0.35
	US	PLACEBO	38	-0.45	-0.34	0.64	0.10
		VEC-162	38	-0.80	-0.45	1.12	0.18
3. MoST	Germany	PLACEBO	4	-0.18	-0.09	0.66	0.33
		VEC-162	4	0.72	0.82	0.84	0.42
	US	PLACEBO	38	0.33	0.22	0.78	0.13
		VEC-162	38	0.48	0.33	0.61	0.10
4. CGIC	Germany	PLACEBO	2	4.50	4.50	0.71	0.50
		VEC-162	4	1.75	1.75	0.29	0.14
	US	PLACEBO	33	3.27	3.50	1.02	0.18
		VEC-162	32	2.73	3.00	1.19	0.21
5. nTST	Germany	PLACEBO	4	-0.12	-0.26	0.68	0.34
		VEC-162	4	0.93	0.96	0.45	0.22
	US	PLACEBO	38	0.37	0.39	0.52	0.08
		VEC-162	38	0.59	0.55	0.99	0.16
6. dTSD	Germany	PLACEBO	4	-0.09	-0.16	1.03	0.51
		VEC-162	4	-0.25	-0.18	0.33	0.17
	US	PLACEBO	38	-0.21	-0.12	0.34	0.06
		VEC-162	38	-0.35	-0.16	0.52	0.08

Source: Reviewer's Analysis
*: For LQ-nTST, MoST and nTST, larger values indicate better outcome; for UQ-dTSD, CGIC and dTSD, smaller values indicate better outcome.

4.1.2 STUDY 3203 (RESET STUDY)

There were 20 patients (10 patients per group) randomized in Study 3203. Due to small sample size, the results of the subgroup analysis for this study aren't presented in this review.

4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

This original NDA submission includes two pivotal efficacy studies, Study 3201 (SET Study) and Study 3203 (RESET Study).

Study 3201 was a multicenter, randomized, double-masked, placebo-controlled, parallel study designed to evaluate the efficacy and safety of 20 mg of tasimelteon versus placebo in patients suffering from Non-24. Eighty-four patients (tasimelteon 42; placebo 42) were randomized to receive tasimelteon (20 mg/day) or placebo. Of these 84 patients who were randomized to study drug, 62 (73.8%) patients completed the Randomization Phase and 22 (26.2%) patients discontinued early. This study was conducted at investigative sites in the US and Germany. The study began with a Pre-Randomization Phase (~5-6 weeks) and was followed by either a Randomization Phase (~26 weeks) or an Open-Label Extension Phase (~26 weeks).

Study 3203 was a multicenter, randomized withdrawal, double-masked, placebo-controlled, parallel group study designed to evaluate the long-term maintenance effect and safety of 20 mg of tasimelteon versus placebo in patients with Non-24. Every patient who enrolled in Study 3203 had previously been screened in 3201. Patients who met the inclusion criteria and who had previously participated in, or were screened for Study 3201, were eligible to participate. The study had 2 phases: a Pre-Randomization Phase (consisting of an Open-label tasimelteon Run-in Phase [~6 weeks] and a τ Estimation Phase [~ 6 weeks]), and a Randomized Withdrawal Phase (~8 weeks). Twenty patients in the US were randomized into the study and they all completed the Randomized Withdrawal Phase.

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the clinical endpoints, instead of a melatonin-based biomarker proposed by the sponsor. The sections below present sponsor's efficacy analyses, summary of important events related to statistical analysis in Study 3201, summary of statistical issues, and this reviewer's efficacy analyses.

5.1.1 SPONSOR'S EFFICACY ANALYSES

Study 3201:

The primary efficacy endpoints were the following:

- The entrainment of the circadian melatonin rhythm as measured by urinary aMT6s (Entrainment is a melatonin-based biomarker and is defined as having a post-baseline τ value less than 24.1 and a 95% CI that included 24.0. τ is circadian period.)

- (Step-down primary endpoint) The Clinical Response rate corresponding to individuals who had both entrainment of the aMT6s rhythm and a score of ≥3 on the Non-24 Clinical Response Scale (N24CRS). N24CRS is a 4-item scale that includes the Lower Quartile-nighttime Total Sleep Time (LQ-nTST), Upper Quartile-daytime Total Sleep Duration (UQ-dTSD), Midpoint of Sleep Time (MoST), and Clinician Global Impression-Change (CGI-C) assessments.

The proportion of Non-24 patients who were entrained after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of Non-24 patients who were entrained after placebo treatment (% difference = 17.4; p = 0.0171). The proportion of patients who were entrained (aMT6s) and had a clinical response rate (N24CRS) ≥3 after tasimelteon treatment during the Randomization Phase was statistically significantly greater than the proportion of patients who were entrained and had a clinical response rate ≥3 after placebo treatment (% difference = 23.7; p = 0.0028).

Study 3203

The primary efficacy endpoint was the proportion of non-entrainment of the circadian melatonin rhythm as measured by urinary aMT6s. The proportion of Non-24 patients who became non-entrained to a 24-hour day after randomization to tasimelteon was statistically significantly less than the proportion of Non-24 patients who became non-entrained after randomization to placebo treatment (% difference = -70.0; p = 0.0026).

5.1.2 Summary of Important Events Related to Statistical Analysis in Study 3201

In the original protocol, the primary endpoint proposed by the sponsor was nTST and the proposed sample size was 160 patients, based on the postulated mean treatment difference of 39 minutes and standard deviation of 66 minutes. In Amendment 6 submitted to the Agency, the sample size was changed from 160 to 100 patients, based on the new postulated mean treatment difference of 30 minutes and standard deviation of 45 minutes. In Amendment 9, the primary endpoint was changed to entrainment and the sample size was reduced to 84 patients. At the time of Amendment 9 (May 21, 2012), 95% of the patients were randomized and 56% of the patients completed the study. Amendment 11 was dated on December 11, 2012 and the trial data was unblinded on December 12, 2012. It is unclear to the Agency how much these changes might have impacted the trial results.

5.1.3 Summary of Statistical Issues

Study 3201

There are three main statistical issues for this study:

Primary Efficacy endpoint(s)

Sponsor's primary efficacy analyses were based on a melatonin-based biomarker endpoint "entrainment". However, throughout the development program for tasimelteon, the Agency repeatedly disagrees that this biomarker is adequate to assess the efficacy for this indication and requires clinically meaningful primary endpoint(s). No agreement was reached between the Agency and the sponsor regarding the primary endpoint(s). The Agency has decided that the

primary efficacy evaluation for tasimelteon should be based on the following clinical endpoints: LQ-nTST, UQ-dTSD, MoST, CGIC, nTST and dTSD.

Analysis Populations

For Study 3201, 84 patients were randomized (n=84). The ITT Population defined by the sponsor included all patients randomized into the study that had τ calculated post-randomization (thereafter referred as Sponsor ITT, n=78). The Analysis Population defined by the sponsor included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization (thereafter referred as Sponsor Analysis Population, n=72). The sponsor's ITT and Analysis Population were selected after the randomization and they are non-randomized subsets. Six patients in the Randomized Population without τ calculated post-randomization were excluded from the Sponsor ITT and six patients in the Sponsor ITT were excluded from the Sponsor Analysis Population due to less than 70% of 1 circadian cycle of nTST data reported during screening and/or post-randomization. Thus, 12 patients in total were excluded from the 84 patients randomized. In sponsor's efficacy analyses, the Sponsor ITT population was utilized for all circadian rhythm related outcomes and the Analysis Population was used for analyses of all other endpoints including the step-down primary and all other efficacy analyses. However, the 84 randomized patients all took study medication and had at least one baseline and postbaseline assessment. Based on the intent-to-treat principal, all the 84 patients should be included in the ITT population (n=84). The ITT population is a randomized population. The efficacy analyses conducted by this reviewer are based on ITT population. In Section 3.2.3.4, this reviewer will present information regarding the 12 patients excluded from the efficacy analyses by the sponsor and explain why these 12 patients should be included in the efficacy analyses.

Analysis Methods

In sponsor's efficacy analyses for the clinical endpoints, ANCOVA model was used which includes baseline value as a covariate and treatment group and pooled sites as factors. However, due to small sample size, non-normal distribution of the data, and some heterogeneity in variances, this reviewer thinks permutation ANCOVA is more appropriate than ANCOVA for analyzing the clinical endpoints. In addition, in sponsor's ANCOVA analysis for Study 3201, study site was included as a factor, but the Study Report shows that the randomization was not stratified by study site. Normally, if the randomization isn't stratified by study site, in order to comply with the trial design, the site is not necessarily included in the analysis model. Therefore, this reviewer thinks permutation ANCOVA without site is more appropriate than ANCOVA with site.

Study 3203

For this study, 20 randomized patients all received the study medication and completed the study, thus all 20 patients were included in the ITT population. The issues related to primary endpoint and analysis method are similar to Study 3201, but please note that for Study 3203, the randomization was not stratified by site and site was not a factor in sponsor's ANCOVA analysis.

5.1.4 REVIEWER'S EFFICACY ANALYSES

For both Study 3201 and Study 3203, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The analyses presented below were conducted by this reviewer and they are post-hoc exploratory analyses without multiplicity adjustment.

In order to visually examine the distribution of the clinical endpoints, this reviewer generated histograms for each of the clinical endpoints in both studies. Please refer to Section 3.2.3.3 for details.

For Study 3201, sponsor's efficacy analyses for clinical endpoints were based on Analysis Population. The Analysis Population included all patients in the Sponsor ITT population that had at least 70% of 1 circadian cycle of nTST data reported during each screening and post-randomization. Based on this definition, 12 patients in the Randomized Population were excluded from the efficacy analysis for clinical endpoints. However, among these 12 patients, 10 patients have more than 50% of 1 circadian cycle of nTST at baseline and 6 patients have more than 50% of 1 circadian cycle of nTST at postbaseline. It seems that the data are not extremely sparse for these 12 patients and this reviewer thinks they should be included in the efficacy analysis. It appears that the inclusion of the 12 patients neither consistently strengthen nor consistently weaken the treatment effect for different clinical endpoints. Please refer to Section 3.2.3.4for details.

This reviewer conducted ANCOVA analysis on ITT for clinical endpoints. In sponsor's ANCOVA analysis for Study 3201, baseline was included as a covariate if applicable and the pooled sites as a factor (variable name: SITEGR3). However, since the randomization wasn't stratified by site, this reviewer thinks it isn't necessary to include site as a factor in the analysis model. Based on the results of ANCOVA analysis without site, it seems that for both studies the nominal p-values were statistically significant or marginally significant for LQ-nTST, UQ-dTSD, MoST, CGIC and dTSD. The nominal p-value for nTST wasn't statistically significant in either of the studies. Please refer to Section 3.2.3.5 for details.

Because of small sample size, non-normal distribution of the clinical endpoints, and some heterogeneity in the variances of the clinical endpoints, this reviewer thinks permutation ANCOVA would be more appropriate than ANCOVA. It appears that the results of permutation ANCOVA without site is fairly close to those of ANCOVA without site. For LQ-nTST, UQ-dTSD, MoST, CGIC, and dTSD, the nominal p-values are statistically significant or marginally significant for both studies. The p-value for nTST isn't statistically significant for either of the studies. Please refer to Section 3.2.3.6 for details.

This reviewer performed subgroup analysis by sex, race, age group (<50 and >=50) and country (Germany and US) for Study 3201. It seems that the point estimates of treatment effects are all in the right direction for sex, age group and country for all clinical endpoints. For race, the point estimates of treatment effect for LQ-nTST, MoST and CGIC aren't in the right direction for non-white patients. However, this doesn't raise concerns since there are only 14 non-white patients in this study. The subgroup analysis for study 3203 wasn't performed due to small sample size (n=20). Please refer to Section 4.1 for details.

5.2 Conclusions and Recommendations

This original NDA submission includes two pivotal efficacy studies, Study 3201 (SET Study) and Study 3203 (RESET Study). For both studies, no agreement was reached between the Agency and the sponsor regarding the primary efficacy endpoint(s), analysis populations and analysis methods. The Agency has decided that the primary efficacy evaluation for tasimelteon should be based on the clinical endpoints. According to this reviewer's post-hoc exploratory analyses without multiplicity adjustment, five out of the six clinical endpoints yield nominal p-values less than 0.05 or approximately 0.05; thus, the two studies appear to suggest that tasimelteon 20 mg may be beneficial for Non-24 Hour Disorder in Totally Blind Individuals based on all the clinical endpoints except for night Total Sleep Time (nTST, the original primary endpoint). However, the results of these two studies should be interpreted with caution, since they are based on post-hoc exploratory analysis without multiplicity adjustment and thus it is unknown whether the overall type I error is properly controlled.

APPEARS THIS WAY ON ORIGINAL

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

/s/

JINGYU J LUAN
11/19/2013

KUN JIN
11/19/2013
I concur with the review,

KOOROS MAHJOOB
11/21/2013
I concur with the review.

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

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDY

NDA:	205677
Drug Name:	TASIMELTEON® 20mg capsules (VEC-162, a melatonin agonist) (Orphan Drug)
Indication:	Treatment of Non-24-Hour Sleep-Wake Disorder in the Totally Blind
Applicant:	Vanda Pharmaceuticals, Inc., Washington, DC CRO: (b) (4)
Date:	Submitted 31 May 2013 Assigned to Reviewer: 26 August 2013
Review Priority:	Priority
Biometrics Division:	Division 6
Statistical Reviewer:	Steve Thomson
Concurring Reviewer:	Team Leader: Karl Lin, Ph. D.
Medical Division:	Neurology Products
Toxicologist:	Reviewer: Melissa Banks-Muckenfuss, Ph.D Team Leader: Lois Freed, Ph.D.
Project Manager:	Cathleen Michaloski, BSN, MPH
Keywords:	Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY 3

1.1. CONCLUSIONS AND RECOMMENDATIONS 3
1.2. BRIEF OVERVIEW OF THE STUDIES 7
1.3. STATISTICAL ISSUES AND FINDINGS 8
1.3.1. Statistical Issues 8
1.3.2. Statistical Findings 14

2. INTRODUCTION 14

2.1. OVERVIEW 14
2.2. DATA SOURCES 14

3. STATISTICAL EVALUATION 14

3.1. EVALUATION OF EFFICACY 14
3.2. EVALUATION OF SAFETY 14
3.2.1 STUDY VEC-162: CARCINOGENICITY STUDY BY ORAL GAVAGE ADMINISTRATION TO CD RATS FOR 104 WEEKS. 14
3.2.2 STUDY VEC-162: CARCINOGENICITY STUDY BY ORAL GAVAGE ADMINISTRATION TO CD-1 MICE FOR 104 WEEKS. 22

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 25

5. SUMMARY AND CONCLUSIONS 26

5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 26
5.2. CONCLUSIONS AND RECOMMENDATIONS 26

APPENDICES: 27

APPENDIX 1. SURVIVAL ANALYSIS 27
APPENDIX 2. FDA POLY-K TUMORIGENICITY ANALYSIS 31
APPENDIX 3. REFERENCES 50

Reference ID: 3401357

1. EXECUTIVE SUMMARY

According to the reports provided by Contract Research Organization, this submission was intended to assess the carcinogenic potential of daily administration of tasimelteon, labeled as VEC-162 in the studies, a melatonin agonist being developed for treatment of sleep disorders, when administered orally by gavage to both rats and CD-1 mice for 104 weeks. Both studies were conducted by [redacted] (b) (4) The description of the studies is taken from the [redacted] (b) (4) final reports.

1.1. Conclusions and Recommendations

This submission summarizes the results from both a standard rat study and a mouse study. The Sponsor's report summarizes the design of the rat study as follows: "The carcinogenic potential of VEC-162 (a melatonin agonist) was assessed over a period of 102 weeks in male and 104 weeks in female Crl:CD® (SD)IGS BR rats by oral gavage administration. Three groups, each comprising 65 males and 65 females, received VEC-162 by gavage at doses of 20, 100 or 250 mg/kg/day. A similarly constituted Control group received the vehicle (100% polyethylene glycol-400) at the same volume-dose." (page 10 of rat report)

Similarly, for the mouse study: "The carcinogenic potential of VEC-162 (a melatonin agonist) was assessed over a period of 104 weeks in Crl: CD-1™ (ICR) BR mice by oral administration. Three groups, each comprising 66 males and 66 females, received VEC-162 by gavage at doses of 30, 100 or 300 mg/kg/day. A similarly constituted Control group received the vehicle (100% polyethylene glycol-400) at the same volume-dose. A further 39 males and 39 females were allocated to each group and were used to provide blood samples for toxicokinetic evaluation." (page 10 of mice report)

As noted in Section 1.3.1.2, the Kaplan-Meier (product-limit) estimate of the survivor function is a valuable graphic to represent survival. Survival curves for each gender in each species are given in Appendix 1. Summary incidence of death tables are presented on pages 18 and 19 and 24 of this report. In male rats the Kaplan-Meier estimated survival curves of the four dose groups show overall lower survival in the low dose group while the medium dose group has generally the highest survival, with survival in the high dose and vehicle crossing, but mostly bounded between these survival curves. This explains the results in the statistical tests of differences in survival among male rats given below:

Table 1. Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0223	0.0082	0.5557	0.3796
No Trend over all four groups	0.1225	0.2021	0.2271	0.2051
No difference between high dose and vehicle	0.4005	0.7710	0.1343	0.3311

Reference ID: 3401357

Thus, the tests of homogeneity in survival over doses in male rats are statistically significant at the usual 0.05 level (Logrank p = 0.0233, Wilcoxon p = 0.0082). However, from the Kaplan-Meier curve, it seems difficult to attribute this result to any particular dose related effect. Rather results are consistent with the lack of evidence for any particular statistically significant dose related trend (Logrank p = 0.1225, Wilcoxon p = 0.2021) and with the lack of any consistent dominance of the survival curves for the high dose and vehicle (Logrank p = 0.4005, Wilcoxon p = 0.7710). In female rats, things are rather different. Whether analyzing trend among all four dose groups or the just the simple comparison of the High dose and vehicle, no tests were statistically significant (i.e., all 6 p ≥ 0.1343). While absence of proof is not proof of absence, the lack of evidence for such differences in survival in female rats is consistent with the hypotheses of no differences or trends.

Table 2. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.6054	0.6502	0.1549	0.0948
No Trend over all four groups	0.7949	0.5793	0.9095	0.9852
No difference between high dose and vehicle	0.4564	0.3617	0.7822	0.6885

In male mice, from the survival curve shown in Figure A.1.3 in Appendix 1, the vehicle eventually slightly dominates the survival curves in the other groups, but this is not sufficient to result in any statistically significant test of homogeneity, dose related trend, or difference between the high dose and control (i.e., all 6 p ≥ 0.3617). From Figure A.1.4, in female mice it does seem that the vehicle, low, and high dose groups are eventually intertwined, but with higher survival than the medium dose group. As the Wilcoxon statistic is more sensitive to later differences, this is sufficient to result in a test of overall homogeneity that is arguable somewhat close to significance (Wilcoxon p = 0.0948), but less so when weighting earlier differences (Logrank p = 0.1549). However, there is no particular evidence of dose related trend or heterogeneity between the high dose and control (all 4 remaining p ≥ 0.6885).

Note that a large number of tumors are typically identified in the analysis of neoplasms, implying a large number of statistical tests. The problem of adjusting for the multiplicity of statistical tests is discussed in Section 1.3.1.4, below. Following the frequentist paradigm, when interpreting significance levels (i.e., p-values), this reviewer would recommend using the Haseman-Lin-Rahman (HLR) rules to adjust for the multiplicity of tests. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate for the joint tests in a two species submission to roughly 10%, one compares the unadjusted significance level of the trend test to 0.005 for common tumors and 0.025 for rare tumors, and the pairwise test between the high dose and control to 0.01 for common tumors and 0.05 for rare tumors. Using these adjustments for other tests, like testing the comparisons between the Low and Medium dose groups versus vehicle can be expected to increase the overall type I error rate to some value above the nominal rough 10%

Reference ID: 3401357

level, possibly considerably higher than the nominal 10% rate. Note that if one tests trend and pairwise differences jointly, a slightly different set of bounds is used.

Table 3 below, shows the tumors that had at least one non-multiplicity adjusted test of trend in rats that was statistically significant at or moderately close to a 0.05 level.

Table 3. Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.7	36.1	49.9	43.2				
HEPATOCELLULAR ADENOMA	2	2	6	8	.0123	.0444	.1802	.6222
Adj. # at Risk	43.7	36.1	50.7	43.2				
Hepatocellular Adenoma/Carcinoma	2	2	8	8	.0145	.0444	.0746	.6222
PANCREAS								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
ISLET CELL ADENOMA	1	5	5	2	.6541	.5000	.1411	.0692
Adj. # at Risk	42.9	36.1	49.3	43.0				
Islet Cell Adenoma/Carcinoma	2	6	7	2	.7843	.6921	.1212	.0878
PITUITARY								
# Evaluated	63	65	65	64				
Adj. # at Risk	47.6	45.6	55.0	46.9				
ADENOMA, PARS DISTALIS	19	28	30	23	.4822	.2361	.1104	.0296
SKIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.1	35.7	50.3	43.4				
FIBROMA	1	0	5	3	.0920	.3080	.1404	1
Adj. # at Risk	43.1	35.7	50.3	43.9				
Fibroma/Sarcoma NOS/Fibrosarcoma	1	0	5	4	.0407	.1800	.1404	1
TESTES								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.3	35.7	49.2	43.2				
INTERSTITIAL (LEYDIG) CELL ADENOMA	3	1	4	6	.0589	.2417	.5735	.9135
Female Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.4	39.6	40.5	40.4				
HEPATOCELLULAR ADENOMA	2	1	6	7	.0087	.0539	.0980	.8511
MAMMARY								
# Evaluated	64	65	64	65				
Adj. # at Risk	47.6	42.4	43.5	44.8				
MAMMARY ADENOCARCINOMA	10	10	12	18	.0146	.0356	.3134	.4864
OVARIES								
# Evaluated	64	65	64	63				
Adj. # at Risk	45.2	39.6	39.3	38.2				
SERTOLIFORM TUBULAR ADENOMA	0	0	0	2	.0546	.2066	.	.
Adj. # at Risk	45.2	39.6	39.4	38.2				
Thecal Cell Tmr/Serto.Tub.Adenoma	0	0	1	2	.0525	.2066	.4643	.
UTERINE CERVIX								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.2	39.6	39.3	39.3				
SQUAMOUS CELL CARCINOMA	0	0	0	2	.0568	.2126	.	.

Reference ID: 3401357

Table 3. (cont.) Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Female Rats (cont.)								
UTERUS								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.8				
Endo. Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722
Adj. # at Risk	45.5	39.6	39.5	39.9				
SQUAMOUS CELL CARCINOMA	1	0	1	3	.0543	.2558	.7160	1
Uterus(w/uterine cervix)								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endometrial Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722
Addj. # at Risk	45.5	39.6	39.5	40.8				
SQUAMOUS CELL CARCINOMA	1	0	1	5	.0059	.0765	.7160	1

All the tumors in male rats listed above would be classified as common (incidence > 1%). Using the Haseman-Lin-Rahman rules described in 1.3.1.5, in male rats no test of trend or pairwise difference between the high dose and control was significant either when testing trend or pairwise difference, although the joint tests of liver tumors would be equivocal. In female rats, in the table above, sertoliform tubular adenoma, plus pooled thecal cell tumor and sertoliform tubular adenoma, both of the ovaries, and squamous cell carcinoma of the uterine cervix would be classified as rare tumors (but not in the uterus or pooled uterus and cervix), the remainder as common. Tumor incidence based on combining the uterus with the uterine cervix only affects the results for squamous cell carcinoma. Going down the table of organ tumor combinations, the test of trend in hepatocellular adenoma of the liver in female rats would be close to the multiplicity adjusted level of statistical significance ($p = 0.0087 \approx 0.005$). The corresponding tests comparing the high dose to vehicle might be classified as close to significant when testing jointly, but not seperately ($p = 0.0539 \approx 0.05$, but > 0.01). The tests of trend in endometrial adenocarcinoma and pooled adenoma and adenocarcinoma of the uterus (and thus the uterus with uterine cervix) would both be categorized as statistically significant ($p = 0.0002$, $0.0005 < 0.005$, respectively), while the corresponding tests between the high dose and vehicle were statistically significant or close ($p = 0.0035 < 0.01 < 0.05$ and $0.0102 \approx 0.01 < 0.05$). Strictly following the HLR rules and specifying that squamous cell carcinoma of the uterus with cervix is a common tumor, we would conclude that both the test of trend and the pairwise comparison are close to statistical significance ($p = 0.0059 \approx 0.005$ amd $0.0765 \approx 0.05$). The test of trend would No other tests of carcinogenicity between dose groups achieved the multiplicity adjusted levels of significance.

Table 4, below, shows similar results in mice. In mice the only test of trend that met the multiplicity adjusted limits for statistical significane was in male mice, namely in adenomas in the caecum or colon or duodenum (i.e the intestine) ($p = 0.0134 < 0.025$), while the corresponding pairwise test was close to significance ($p = 0.1025 \approx 0.10$), for testing hypotheses jointly (please see Section 1.3.1.5). Note, however, that there is a bit of a conundrum in that

Reference ID: 3401357

strictly following the vehicle incidence to determine whether or not a tumor can be classified as common or rare, the test of trend in the more complete intestine (adding the jejunum) and pooling adenoma and adenocarcinoma would not be statistically significant (p = 0.0114 > 0.005).

Table 4. Potentially Statistically Significant Neoplasms in Mice

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Mice								
Caecum+Colon+Duodenum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
Adenoma	0	0	0	3	.0134	.1025	.	.
Caecum+Colon+Duodenum+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.6	45.8	46.5	45.4				
Adenoma/Adenocarcinoma	2	0	0	5	.0114	.1764	1	1
DUODENUM								
# Evaluated	66	65	65	66				
Adj. # at Risk	50.1	44.9	46.5	45.1				
ADENOMA	0	0	0	2	.0582	.2217	.	.
Female Mice								
UTERINE CERVIX								
# Evaluated	66	64	65	66				
Adj. # at Risk	48.6	44.3	39.8	48.1				
ENDOMETRIAL POLYP	0	0	1	2	.0652	.2474	.4483	.

Complete results of statistical poly-k tests of tumor trend and differences between dose groups are given in Tables A.2.3 through A.2.6 in Appendix 2.

1.2. Brief Overview of the Studies

Two studies were submitted:

VEC-162: Carcinogenicity Study by Oral Gavage Administration to CD Rats for 104 Weeks,

and,

VEC-162: Carcinogenicity Study by Oral Gavage Administration to CD-1 Mice for 104 Weeks.

These studies were designed to assess the carcinogenic potential of daily administration of VEC-162, a melatonin agonist being developed for treatment of sleep disorders, when administered orally by gavage to both rats and CD-1 mice for about 104 weeks. The dose groups in each study were labeled in this report as the Low, Medium, and High dose groups, respectively, plus the Vehicle control group.

Reference ID: 3401357

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

In this section, several issues, typical of statistical analyses of these studies, are considered. These issues include comments on the multiple housing of animals, details of the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Multiple Housing and Dosing of Animals:

The Sponsor's report indicates that unless reduced by mortality or isolation, study animals were housed 5 per cage in the main rat study and 3 per cage in the main mice study. Social interaction is important for the welfare of the animals. However, an argument could be made that housing animals together should be treated as part of the treatment of the experiment, and thus the appropriate unit of analysis would be the cage of the animals, not the individual animal. This would effectively reduce the study sample size and thus decrease power.

Further, carcinogenicity tendencies could be communicated across animals in that competition for food, fighting, or other within cage effects could cause positive or negative correlations in response. When animals in a cage are dosed together or at the same time, variations in dosing across occasions might also induce positive or negative correlations. Thus, it is possible that within treatment estimated variances may be too large or too small, resulting in conservative or anti-conservative tests (in terms of Type I error). Unless it has been clearly shown that tumor incidence is independent of cage, from a purely statistical analysis point of view, this reviewer would generally recommend single housing and dosing of animals.

1.3.1.2. Survival Analysis:

The survival analyses in rats presented here are based on both the log rank test and the Wilcoxon test comparing survival curves. The log rank tests tend to put higher weight on later events, while the Wilcoxon test tends to weight events more equally, and thus is more sensitive to earlier differences in survival. The logrank test is most powerful when the survival curves track each other, and thus the hazards, i.e., the conditional probability of the event in the next infinitesimal interval, would be roughly proportional. Note the logrank test seems to be the test usually recommended by statisticians, and is the test used by the Sponsor. Both tests are used in the FDA analysis of mortality. Appendix 1 reviews the specific FDA animal survival analyses in more detail. The results of the Sponsor's analyses are summarized in Sections 3.2.1.1. and 3.2.2.1.

Arguably the best pictorial representation of survival is a plot of the survival function, S(t), defined as the proportion of subjects surviving to at least time t (i.e. greater than or equal to time t). A natural estimate of this is defined for event times $t1 < t2 < t3 < \ldots < tN$, with dj deaths in the nj animals at risk in time tj (animals that die or leave the study earlier are not at risk). Then for times j=1 to N, the Kaplan-Meier estimated survival curve is defined as the cumulative product at the nth point as $S(tn) = \prod (1 - tj/nj)$ for $j=1,\ldots,n \leq N$.

Reference ID: 3401357

1.3.1.3. Multiplicity of Tests on Survival:

Using both the logrank and Wilcoxon tests, for each gender in each species there are 6 tests of survival differences. Assuming tests were performed at the usual 0.05 level, and the tests were stochastically independent, but there were actually absolutely no differences in survival across groups (so one would hope no tests would be statistically significant), the probability of at least one statistically significant result in each species by gender combination was about 0.265. These bounds assume that the tests are independent, which they clearly are not, but these values can give some idea of the possible price paid for the multiplicity of hypothesis tests in the statistical frequentist paradigm.

1.3.1.4. Tests on Neoplasms:

The data sets requested for the analysis of rodent carcinogenicity studies are supposed to include a record for each animal organ combination that was not evaluated. If a number of the animals are not examined, but the proportions of animals showing the tumor under study in each treatment group is roughly the same as in the subset of animals actually reported the calculated p-values will generally be too large, i.e., results will be less statistically significant than they should be, possibly much less. If we can assume the process that determines whether or not a tumor is analyzed in each specific tumor is random, it is perhaps appropriate to consider such endpoints to be both analyzed AND have the tumor.

Ignoring these possible problems, the Sponsor's analyses of tumorigenicity are based on Peto et al (1966) methods, based on logrank for fatal and mortality independent tumors and a Mantel-Haenzel analysis of incidental tumors stratified by time of detection. Among other problems, this Peto style analysis requires accurate determination of whether a tumor is fatal or incidental.

The FDA analysis is based on a modification of the Cochran-Armitage test of trend in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). Inspecting a large number of studies, Bailer and Portier noted that survival time seemed to fit a Weibull probability distribution, generally with a shape parameter of between 1 and 5, with 3 a typical value. With t_{max} denoting the maximal time to terminal sacrifice and t_{obs} the time to detection of the tumor in the animal, they proposed weighting the animal by $(t_{obs}/t_{max})^k$, so that an animal that survives for say 52 weeks in 104 week study without the tumor being analyzed is counted as $(1/2)^k$ of an animal in the risk set for that tumor. For k = 3, that means that particular animal would count as 1/8 of an animal. Further, the k = 3 specification seems to represent tumor incidence where some animals are perhaps more sensitive and respond earlier to the insult than the remining animals. Under this structure time to incidence would tend to follow a cubic expression. Thus an animal with the specific tumor being studied or who survives to terminal sacrifice without the tumor will be given a weight of 1 when counting the number of animals at risk. However, animals that die early without the tumor are down weighted when counting the number of animals in the risk set for that specific tumor. With differential mortality, this can mean a substantial reduction in the size of that risk set. Note this seems to be an appropriate adjustment

Reference ID: 3401357

for dose groups that are terminated early. The report of the Society of Toxicological Pathology "town hall" meeting in June 2001 recommeded the use of this poly-k modification of the so-called Cochran-Armitage tests of trend over the corresponding Peto tests used by the Sponsor.

The computed significance levels are based on small sample exact permutation tests of tumor incidence. In the tumor incidence tables the effective size of the risk set for each tumor is listed in the row labeled "Adjusted # at risk ", and seems to be a more appropriate denominator when comparing incidence rates than the simple unadjusted number evaluated.

1.3.1.5. Multiplicity of Tests on Neoplasms:

Testing dose related treatment differences for each species by gender by organ by tumor combination involves a large number of comparisons. Current FDA practice is based on the Haseman-Lin-Rahman multiplicity adjustments.

The Haseman-Lin-Rahman rules are based on the original multiplicity adjustment of Haseman (1983) and extended by Lin and Rahman with various simulations. Based on his extensive experience with such analyses, for pairwise tests in a two species study comparing control to the High dose group, Haseman (1983) claimed that for a roughly 0.10 (10%) overall false positive error rate, rare tumors should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Lin & Rahman (1998) proposed a further p-value adjustment for tests of trend. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of trend, rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level. Other specifications are presented in the Table 4 below. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

The proposed Haseman-Lin-Rahman bounds are taken from *Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals*, (HHS, 2013). The bounds on the right in table 5, below, are grouped so that the last four columns correspond to testing either trend or pairwise comparison between the high dose and control seperately. The previous four columns (columns 2-5), correspond to testing both overall trend and pairwise tests between the high dose and control together. Within each group there is a column giving the corresponding bounds for a two species study and another column for a one species study. In this analysis we follow the usual practice of testing parameters separately, so the bounds in the leftmost column are used. The observed tumor incidence in the vehicle group is used to decide if a tumor is classified rare or common.

Reference ID: 3401357

Table 5. Recommended Multiplicity Adjusted Bounds on Significance Levels

	Testing trend or pairwise (not both)				Joint testing of trend and pairwise			
	Two Species		One Species		Two Species		One Species	
	Trend	Pairwise	Trend	Pairwise	Trend	Pairwise	Trend	Pairwise
Common Tumor	0.005	0.01	0.01	0.025	0.005	0.05	0.01	0.05
Rare Tumor	0.025	0.05	0.05	0.10	0.025	0.10	0.05	0.10

The significance levels of the pairwise tests between the vehicle control, and Low and Medium dose groups are also provided in the tumor analysis tables below. Following the HLR rules, adding these comparisons can be expected to increase the overall type I error rate to some level above the usual rough 10% level, possibly considerably larger. Again, because of the possibility of genetic drift and for convenience the vehicle group is used to determine if the tumor is classified as rare or common.

1.3.1.6. Validity of the Designs:

When determining the validity of designs there are two key points:
1) adequate drug exposure
2) tumor challenge to the tested animals.

1) is related to whether or not sufficient animals survived long enough to be at risk of forming late-developing tumors and 2) is related to the Maximum Tolerated Dose (MTD), designed to achieve the greatest likelihood of tumorigenicity.

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals (i.e., 50%), between weeks 80-90 of a two-year study may be considered a sufficient number of survivors as well as one measure of adequate exposure. From tables 14 on page 18, 15 on page 19, and 20 and 21 on page 24, as a percentage of the High dose group animals that survived to week 91, this criterion does seem to be not quite met in female rats (Males: 58.5% and Females: 44.6%), but is exceded in male rats and mice (Males: 60.6% and Females: 65.2%).

The mean weight values used to derive differences and ratios in the following tables were taken directly from the Sponsor's reports (both labeled Table 4, Body Weight group mean values (g), pages 71-78 in rats and 53-60 in mice). The mean change from baseline is that reported by the Sponsor and is calculated from the weight changes of individual animals surviving the specified period.

Reference ID: 3401357

Table 6. Mean Weights and Changes (in g) in Male Rats

Dose Group	Dose mg/kg/day	Week 0	Week 102[1]	Change from Baseline	% change relative to vehicle
1. Vehicle	0	169	737	572	
2. Low	20	171	708	545	95%
3. Medium	100	167	710	545	95%
4. High	250	172	682	513	90%

[1]Control group reached 20 survivors at Week 102, so all groups were sacrificed.

Table 7. Mean Weights and Changes (in g) in Female Rats

Dose Group	Dose mg/kg/day	Week 0	Week 104	Change from Baseline	% change relative to vehicle
1. Vehicle	0	142	466	329	
2. Low	20	149	503	354	108%
3. Medium	100	144	453	309	94%
4. High	250[1]	144	445	297	90%

[1]Dosing stopped at Week 101.

Table 8. Mean Weights and Changes (in g) in Male Mice

Dose Group	Dose mg/kg/day	Week 0	Week 104	Change from Baseline	% change relative to vehicle
1. Vehicle	0	34.2	49.2	14.9	
2. Low	30	34.2	48.8	13.9	93%
3. Medium	100	33.6	46.7	12.7	85%
4. High	300	34.3	44.9	11.1	74%

Table 9. Mean Weights and Changes (in g) in Female Mice

Dose Group	Dose mg/kg/day	Week 0	Week 71	Change from Baseline	% change relative to vehicle
1. Vehicle	0	25.7	39.1	13.7	
2. Low	30	25.7	40.0	14.6	107%
3. Medium	100	25.9	40.6	15.2	111%
4. High	300	25.9	40.0	14.0	102%

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag et al. (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " From Table 8 above, it seems that the weight criterion is exceded in the high dose male mice, as well as possibly in the medium

Reference ID: 3401357

dose, but as indicated in Tables 6, 7, and 9, not in rats or female mice. Further, table 9 in female mice indicates thar all three treatment groups gained weight relative to the vehicle control. This may be evidence that the dose was under the MTD.

The Sponsor reports than in the rat study food consumption was not affected by treatment, while in mice food consumption was low from Week 16 for males and females receiving the High dose, 300 mg/kg/day.

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. This suggests that a useful way to assess whether or not the MTD was achieved is to measure early mortality not associated with any identified tumor. If this is high in the higher dose groups it suggests that animals tend to die before having time to develop tumors. Table 10, below, displays the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors (i.e., the "Event"):

Table 10. Natural Death with No Identified Tumor in Rats (Male/Female)

	1.Vehicle	2. Low	3. Medium	4. High
Males Event	14	22	13	16
No event	51	43	52	49
Females Event	3	5	5	9
No event	62	60	60	56

It is apparent that there is no dose related trend in dying without tumor in male rats. Further, there is no particular evidence of the more general hypothesis of heterogeneity in this event among the dose groups (chi-square p = 0.2615, Fisher exact p = 0.2741). Due the relative rarity of events, there is no overwhelming evidence of heterogeneity in female rats (chi-square p = 0.2869, Fisher exact p = 0.3304). However the more powerful test of dose related trend, i.e. the Cochran-Armitage test is statistically significant (p = 0.0324), although not overwhelmingly so.

Table 11. Natural Death with No Identified Tumor in Mice (Male/Female)

	1.Vehicle	2. Low	3. Medium	4. High
Males Event	16	15	16	16
No event	50	51	50	50
Females Event	8	11	19	13
No event	58	55	47	53

It is clear that no statistical test of treatment group differences is necessary in male mice. However, for completeness, the tests of heterogeneity over dose group in the event are highly not statistically significant (chi-square p = 0.9959, Fisher exact p = 1.0). There is more evidence of a dose related effect in female mice, although results are not statistically significant at the usual 0.05 level (chi-square p = 0.0981, Fisher exact p = 0.1068, Cochrane-Armitage p = 0.1940). Like the other observations above, this requires the expertise of the toxicologist, but these tests

Reference ID: 3401357

may provide evidence that the MTD was not exceeded in male rats or mice, but is a bit more debatable in in female rats, and may not have even been achieved in female mice.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

Results from studies in Crl:CD® (SD)IGS BR rats and Crl: CD-1™ (ICR) BR mice were submitted to assess the carcinogenic potential of tasimelteon.

2.2. Data Sources

Two SAS data sets, both labeled tumor.sas7bdat, were translated from SAS transport files, both labeled tumor.xpt in each study. These followed the usual Biostatistics requested data format. The toxicologist provided a reasonably extensive guide for combining organs and especially associated tumors.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

More detailed results on the study are presented below.

3.2.1 VEC-162: Carcinogenicity Study by Oral Gavage Administration to CD Rats for 104 Weeks.

STUDY DURATION: Male rats 102 weeks, Females 104 weeks.
DOSING STARTING DATE: 10 April 2006.
TERMINAL SACRIFICE: Males Week 102, Females Week 104.
NECROPSY COMPLETED: Males 25-27 March 2008
Females: 7-9 April 2008
STUDY ENDING DATE (Final Report dated): 27 August 2009.
RAT STRAIN: Crl:CD® (SD)IGS BR Rats.
ROUTE: Daily Oral gavage

Reference ID: 3401357

Animals were dosed once daily by oral gavage. Gross aspects of the study designs for the main study animals are summarized in Table 12 be1ow:

Table 12. Design of Rat Study (dose volume 2.5 mL/kg)

Treatment Group	# Main study animals (# TK[1] animals)/gender	Nominal Dose (mg/kg/day)	Nominal Dosing Concentra-tion (mg/mL)
1. Vehicle[2]	65 (10)	0	0
2. Low	65 (10)	20	8
3. Medium	65 (10)	100	40
4. High	65 (10)	250	100

[1] Toxicokinetic phase animals began dosing during Week 1 of the carcinogenicity phase, were sampled in Weeks 6 and 26 and then terminated. In Week 96, samples were also taken from three male and three female Main Study animals with the lowest animal number.
[2] 100% polyethylene glycol-400 (PEG400).

The Sponsor summarizes study conduct as follows: "The carcinogenic potential of VEC-162 (a melatonin agonist) was assessed over a period of 102 weeks in male and 104 weeks in female Crl:CD® (SD)IGS BR rats by oral gavage administration. Three groups, each comprising 65 males and 65 females, received VEC-162 by gavage at doses of 20, 100 or 250 mg/kg/day. A similarly constituted Control group received the vehicle (100% polyethylene glycol-400) at the same volume-dose." (page 10 of rat report)

"For males, the number of survivors in Group 1 (Control) fell to 20 during Week 102. In accordance with the FDA sacrifice criteria, the entire sex was terminated in Week 103. The duration of treatment for males is therefore reported as 102 weeks.

"For females, the number of survivors in Group 4 (250 mg/kg/day) fell to 20 in Week 101. In accordance to the FDA sacrifice criteria, dosing was suspended for this group/sex for the remainder of the study. Females in Groups 1, 2 and 3 completed the scheduled 104 weeks of treatment. The duration of treatment for females is therefore reported as 104 weeks." (page 13 of rat report)

After acclimation, animals were approximately five to six weeks old at first and their bodyweights were reported to be in the range of 136 to 211g for males and 109 to 174g for females. During the study animals were housed in groups of five animals of the same sex in the main study. Food and water were available ad libitum. The Sponsor states that detailed physical examinations were made on all animals twice weekly for the first month, then weekly for the next three months, and then biweekly and finally monthly. Body weights were recorded weekly for the first 16 weeks, beginning approximately one week before initiation of dosing, and every 4 weeks thereafter. Group mean weight changes were calculated from the weight changes of individual animals surviving the specified period.

Reference ID: 3401357

Dosing was justified as follows: "The doses used in this study (0, 20, 100 and 250 mg/kg/day) were selected by the Sponsor with reference to previous work with this compound performed based on findings from oneand six-month toxicity studies. At the highest dose of 500 mg/kg/day in the six month study, one female was killed in moribund condition after three doses due to severe clinical signs, including seizures and laboured respiration. In addition, the incidence and severity of progressive nephropathy were increased at 500 mg/kg/day. In the one-month toxicity study, doses of 100 or 400 mg/kg/day caused hyaline droplet nephropathy. After a one-month recovery period, the kidney lesions were still detectable in males. Based on these findings, 250 mg/kg/day was selected as the high dose for this carcinogenicity study. The low and mid doses (20 and 100 mg/kg/day) were selected based on appropriateness of the dosing intervals to establish a dose-response." (page 12 of rat report)

3.2.1.1 Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigencity in rats.

Survival analysis:

The Sponsor summarized mortality results as follows: "A total of 162 males and 180 females died or were killed prematurely during the treatment period. In Week 101 the number of female survivors in the high dose group (250 mg/kg/day) fell to 20 and dosing was suspended for this group/sex until the end of the treatment period. Females of the other groups completed the scheduled 104-week treatment period. In Week 102 the number of male survivors in the control group fell to 20 and, consequently, the entire sex was terminated early.

"The group distribution of the deaths (i.e. up to Week 102 in males and Week 104 in females) is presented below and there was no treatment-related effect on mortality." (page 36 of report)

Table 13. Sponsor Table: Group distributions of mortality

Group/sex	**1M**	**2M**	**3M**	**4M**	**1F**	**2F**	**3F**	**4F**
Dose (mg/kg/day)	**0**	**20**	**100**	**250**	**0**	**20**	**100**	**250**
Premature deaths	45	46	33	38	43	45	44	48
Percentage survival	31	29	49	42	34	31	32	26

"Statistical analysis revealed that the [logrank] trend test was not statistically significant when all groups were included in the analysis (p=0.126 and 0.232 for males and females, respectively). Only the pair-wise comparison of mortality in males receiving100 mg/kg/day was statistically significant (p=0.035)." (page 36 of rat report) Note that this comparison corresponds to a intermediate dose group and was not tested in the FDA analysis.

Tumorigenicity analysis:

The Sponsor used a standard Peto analysis as described in Section 1.3.1.4. Results are summarized below:

Reference ID: 3401357

"Males
"Liver …
For benign hepatocellular adenoma, the trend test was statistically significant when all groups were included in the analysis (p=0.010). Upon exclusion of the 250 mg/kg/day treated group the trend test was no longer significant (p=0.082). The pairwise comparison of the control group with the 250 mg/kg/day treated group was statistically significant (p=0.045).

"For benign hepatocellular adenoma and malignant hepatocellular combined, the trend test was statistically significant when all groups were included in the analysis (p=0.011). Upon exclusion of the 250 mg/kg/day treated group the trend test was still significant (p=0.018). The pairwise comparison of the control group with the 250 mg/kg/day treated group was statistically significant (p=0.045).

"Pituitary (pars distalis) ….
For benign adenoma, the trend test was not statistically significant when all groups were included in the analysis (p=0.753). The pairwise comparison of the control group with the 20 mg/kg/day treated group was statistically significant (p=0.017).

"Females
"Liver …
For benign hepatocellular adenoma, the trend test was statistically significant when all groups were included in the analysis (p=0.005). Upon exclusion of the 240 mg/kg/day treated group the trend test was still significant (p=0.017). None of the pairwise tests were statistically significant.

"Mammary areas …
For malignant adenocarcinoma, the trend test was statistically significant when all groups were included in the analysis (p=0.008). Upon exclusion of the 240 mg/kg/day treated group the trend test was no longer significant (p=0.221). The pairwise comparison of the control group with the 240 mg/kg/day treated group was statistically significant (p=0.020).

"Ovaries …
For benign sertoliform tubular adenoma, the trend test was statistically significant when all groups were included in the analysis (p=0.039). None of the pairwise tests were statistically significant.

"Uterus …
For malignant endometrial adenocarcinoma, the trend test was statistically significant when all groups were included in the analysis (p<0.001). Upon exclusion of the 240 mg/kg/day treated group the trend test was no longer significant (p=0.537). The pairwise comparison of the control group with the 240 mg/kg/day treated group was statistically significant (p=0.003).

Reference ID: 3401357

"For benign endometrial adenoma and malignant endometrial adenocarcinoma combined, the trend test was statistically significant when all groups were included in the analysis ($p<0.001$). Upon exclusion of the 240 mg/kg/day treated group the trend test was no longer significant ($p=0.670$). The pairwise comparison of the control group with the 240 mg/kg/day treated group was statistically significant ($p=0.008$).

"For malignant squamous cell carcinoma, the trend test was statistically significant when all groups were included in the analysis ($p=0.041$). Upon exclusion of the 240 mg/kg/day treated group the trend test was no longer significant ($p=0.499$). None of the pairwise tests were statistically significant.

"Uterine cervix …
For malignant squamous cell carcinoma, the trend test was statistically significant when all groups were included in the analysis ($p=0.046$). None of the pairwise tests were statistically significant." (pages 427-428 of rat report)

3.2.1.2 FDA Reviewer's Results

This section will present the current Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 14 for male rats, Table 15 for female rats) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval.

Table 14. Summary of Male Rats Mortality (dose/kg/day)

Period	1.Vehicle	2.Low	3.Medium	4.High
0-52	3/65[1] 95.4%[2]	12/65 81.5%	3/65 95.4%	8/65 87.7%
53-78	15/62 72.3%	16/53 56.9%	7/62 84.6%	12/57 69.2%
79-91	15/47 49.2%	11/37 40.0%	13/55 64.6%	7/45 58.5%
92-104	12/32 30.8%	7/26 29.2%	10/42 49.2%	11/38 41.5%
Terminal 105	20	19	32	27

[1] number deaths / number at risk
[2] per cent survival to end of period.

Reference ID: 3401357

Table 15. Summary of Female Rats Mortality (dose/kg/day)

Period	1.Vehicle	2.Low	3.Medium	4.High
0-52	2/65[1] 96.9%[2]	7/65 89.2%	9/65 86.2%	7/65 89.2%
53-78	11/63 80.0%	15/58 66.2%	11/56 69.2%	13/58 69.2%
79-91	11/52 63.1%	10/43 50.8%	9/45 55.4%	16/45 44.6%
92-102	19/41 33.8%	13/33 30.8%	15/36 32.3%	12/29 26.2%
Terminal 102-104	22	20	21	17

[1] number deaths / number at risk
[2] per cent survival to end of period.

As described in Appendix 1, the Kaplan-Meier (product-limit) estimate of the survivor function is probably the best figure to represent survival. Curves for each gender in each species are given in Appendix 1. Summary incidence of death tables are presented on pages 18 and 24 of this report. In male rats the Kaplan-Meier estimated survival curves of the four dose groups show overall lower survival in the low dose group while the medium dose group has generally the highest survival, with survival in the high dose and vehicle crossing, but mostly bounded between these survival curves.

These descriptions explain the results in the tests of differences in survival given below:

Table 16. Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0223	0.0082	0.5557	0.3796
No Trend over all four groups	0.1225	0.2021	0.2271	0.2051
No difference between high dose and vehicle	0.4005	0.7710	0.1343	0.3311

The tests of homogeneity in survival over doses in male rats are statistically significant at the usual 0.05 level (Logrank $p = 0.0233$, Wilcoxon $p = 0.0082$). From the Kaplan-Meier curve, it seems difficult to attribute this result to any particular dose related effect, but does correspond to treatment differences no clearly related to dose. This seems to be consistent with the lack of any statistically significant dose related trend (Logrank $p = 0.1225$, Wilcoxon $p = 0.2021$) or any significant difference between the the high dose and vehicle (Logrank $p = 0.4005$, Wilcoxon $p = 0.7710$). In female rats, when dealing with all four dose groups things are quite different. Whether analyzing among all four dose groups or the comparison of the High dose and vehicle were not statistically significance (i.e., all 6 $p \geq 0.1343$). While absence of proof is not proof of absence, the lack of evidence for such differences in survival is consistent with the hypotheses of no differences or trends.

Reference ID: 3401357

Tumorigenicity analysis:

Table 17 below, a repeat of Table 3 above, and a slight reduction in Table A.2.1 below, shows those tumors that had at least one non-multiplicity adjusted test that was statistically significant at a about a 0.05 level.

Table 17. Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Level ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.7	36.1	49.9	43.2				
HEPATOCELLULAR ADENOMA	2	2	6	8	.0123	.0444	.1802	.6222
Adj. # at Risk	43.7	36.1	50.7	43.2				
Hepatocellular Adenoma/Carcinoma	2	2	8	8	.0145	.0444	.0746	.6222
PANCREAS								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
ISLET CELL ADENOMA	1	5	5	2	.6541	.5000	.1411	.0692
PITUITARY								
# Evaluated	63	65	65	64				
Adj. # at Risk	47.6	45.6	55.0	46.9				
ADENOMA, PARS DISTALIS	19	28	30	23	.4822	.2361	.1104	.0296
SKIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.1	35.7	50.3	43.9				
Fibroma/Sarcoma NOS/Fibrosarcoma	1	0	5	4	.0407	.1800	.1404	1
Female Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.4	39.6	40.5	40.4				
HEPATOCELLULAR ADENOMA	2	1	6	7	.0087	.0539	.0980	.8511
MAMMARY								
# Evaluated	64	65	64	65				
Adj. # at Risk	47.6	42.4	43.5	44.8				
MAMMARY ADENOCARCINOMA	10	10	12	18	.0146	.0356	.3134	.4864
OVARIES								
# Evaluated	64	65	64	63				
Adj. # at Risk	45.2	39.6	39.3	38.2				
SERTOLIFORM TUBULAR ADENOMA	0	0	0	2	.0546	.2066	.	.
Adj. # at Risk	45.2	39.6	39.4	38.2				
Thecal Cell Tmr/Serto.Tub.Adenoma	0	0	1	2	.0525	.2066	.4643	.
UTERINE CERVIX								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.2	39.6	39.3	39.3				
SQUAMOUS CELL CARCINOMA	0	0	0	2	.0568	.2126	.	.
UTERUS								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endo. Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722

Reference ID: 3401357

Table 17. (cont.) Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Level ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Female Rats (cont.)								
Uterus(w/uterine cervix)								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endometrial Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722
Addj. # at Risk	45.5	39.6	39.5	40.8				
SQUAMOUS CELL CARCINOMA	1	0	1	5	.0059	.0765	.7160	1

All the tumors in male rats listed above would be classified as common (incidence > 1%). Using the Haseman-Lin-Rahman rules described in 1.3.1.5, in male rats no test of trend or pairwise difference between the high dose and control was significant either when testing trend or pairwise difference, although the joint tests of liver tumors would be equivocal. In female rats, in the table above, sertoliform tubular adenoma, plus pooled thecal cell tumor and sertoliform tubular adenoma, both of the ovaries, and squamous cell carcinoma of the uterine cervix would be classified as rare tumors (but not in the uterus or pooled uterus and cervix), the remainder as common. Tumor incidence based on combining the uterus with the uterine cervix only affects the results for squamous cell carcinoma. Going down the table of organ tumor combinations, the test of trend in hepatocellular adenoma of the liver in female rats would be close to the multiplicity adjusted level of statistical significance ($p = 0.0087 \approx 0.005$). The corresponding tests comparing the high dose to vehicle might be classified as close to significant when testing jointly, but not seperately ($p = 0.0539 \approx 0.05$, but > 0.01). The tests of trend in endometrial adenocarcinoma and pooled adenoma and adenocarcinoma of the uterus (and thus the uterus with uterine cervix) would both be categorized as statistically significant ($p = 0.0002$, $0.0005 < 0.005$, respectively), while the corresponding tests between the high dose and vehicle were statistically significant or close ($p = 0.0035 < 0.01 < 0.05$ and $0.0102 \approx 0.01 < 0.05$). Strictly following the HLR rules and specifying that squamous cell carcinoma of the uterus with cervix is a common tumor, we would conclude that both the test of trend and the pairwise comparison are close to statistical significance ($p = 0.0059 \approx 0.005$ amd $0.0765 \approx 0.05$). The test of trend would No other tests of carcinogenicity between dose groups achieved the multiplicity adjusted levels of significance.

Complete results of statistical poly-k tests of tumor trend and differences between dose groups in male rats are given in Tables A.2.3 and A.2.4, in Appendix 1.

Reference ID: 3401357

3.2.2 VEC-162: Carcinogenicity Study by Oral Gavage Administration to CD-1 Mice for 104 Weeks.

STUDY DURATION: 104 weeks.
DOSING STARTING DATE: 27 April 2006.
TERMINAL SACRIFICE: Week 104.
NECROPSY COMPLETED: 2 May 2008
STUDY ENDING DATE (Final Report dated): 15 July 2009.
MICE STRAIN: Crl: CD-1™ (ICR) BR Mice.
ROUTE: Daily oral gavage

The drug vehicle is 100% polyethylene glycol-400 (PEG400). Animals were dosed once daily by oral gavage. Gross aspects of the study designs for the main study animals are summarized in Table 18 be1ow:

Table 18. Design of Mice Study (dose volume 4 mL/kg)

Treatment Group	# Main study animals (# TK[1] animals)/gender	Nominal Dose (mg/kg/day)	Nominal Dosing Concentra-tion (mg/mL)
1. Vehicle[2]	66 (39)	0	0
2. Low	66 (39)	30	7.5
3. Medium	66 (39)	100	25
4. High	66 (39)	300	75

[1] Toxicokinetic phase animals began dosing during Week 1, were samples during Weeks 4 and 26, then terminated during Week 26.
[2] 100% polyethylene glycol-400.

The Sponsor summarizes the study conduct as follows: "The carcinogenic potential of VEC-162 (a melatonin agonist) was assessed over a period of 104 weeks in Crl: CD-1™ (ICR) BR mice by oral administration. Three groups, each comprising 66 males and 66 females, received VEC-162 by gavage at doses of 30, 100 or 300 mg/kg/day. A similarly constituted Control group received the vehicle (100% polyethylene glycol-400) at the same volume-dose. A further 39 males and 39 females were allocated to each group and were used to provide blood samples for toxicokinetic evaluation." (page 10 of mice report)

Animals were approximately five to six weeks old at first dosing. During the study animals were housed in groups of three animals of the same sex. Food and water were available ad libitum. The Sponsor states that detailed physical examinations were made on all animals "daily during the first week of treatment, twice weekly during Weeks 2 to 4 (middle and end of each week), once weekly during Weeks 5 to 13, once every two weeks during Weeks 14 to 52 and once every four weeksthereafter." (page 20 of mice report)

Reference ID: 3401357

3.2.1.1 Sponsor's Results and Conclusions

This section will present a summary of the Sponsor's analysis on survivability and tumorigencity in rats.

Survival analysis:

The Sponsor summarized mortality results as follows:

"A total of 133 males and 162 females died or were killed prematurely during the treatment period. The group distribution of the deaths is presented below and there was no treatment related effect on mortality." (page 33)

Table 19. Sponsor Table: Group distribution of mortality

Group/sex	**1M**	**2M**	**3M**	**4M**	**1F**	**2F**	**3F**	**4F**
Dose (mg/kg/day)	**0**	**30**	**100**	**300**	**0**	**30**	**100**	**300**
Premature deaths	30	38	32	33	37	41	47	38
Percentage survival	55	42	52	50	44	38	30	42

"Statistical analysis revealed that the trend test was not statistically significant when all groups were included in the analysis (p=0.802 and 0.897 for males and females, respectively). None of the pair-wise comparisons was statistically significant." (page 33 of report) Note that this table is consistent with tables 20 and 21 reported in the FDA analysis in Section 3.2.1.2, below.

Tumorigenicity analysis:

"There were no neoplastic changes that were attributed to treatment." (page 36 of mouse report) At least to the extent that there were no statistically significant, multiplicity adjusted results, the statisistical analysis below is consistent with this conclusion.

3.2.1.2 FDA Reviewer's Results

This section will present the current Agency findings on survival and tumorigenicity in male and female rats.

Survival analysis:

Kaplan-Meier plots comparing treatment groups in both studies are given in Appendix 1, along with more details of the analysis. The following tables (Table 20 for male mice, Table 21 for female mice) summarize the mortality results for the dose groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent survived at the end of the interval.

Reference ID: 3401357

Table 20. Summary of Male Mice Mortality (dose/kg/day)

Period	1.Vehicle	2.Low	3.Medium	4.High
0-52	6/66[1] 90.9%[2]	4/66 93.9%	9/66 86.4%	8/66 87.9%
53-78	5/60 83.3%	13/62 74.2%	10/57 71.2%	10/58 72.7%
79-91	8/55 71.2%	7/49 63.6%	4/47 65.2%	8/48 60.6%
92-104	11/47 54.5%	14/42 42.4%	9/43 51.5%	7/40 50.0%
Terminal 105	36	28	34	33

[1] number deaths / number at risk
[2] per cent survival to end of period.

Table 21. Summary of Female Mice Mortality (dose/kg/day)

Period	1.Vehicle	2.Low	3.Medium	4.High
0-52	3/66[1] 95.5%[2]	3/66 95.5%	9/66 86.4%	2/66 97.0%
53-78	9/63 81.8%	12/63 77.3%	13/57 66.7%	17/64 71.2%
79-91	9/54 68.2%	10/51 62.1%	14/44 45.5%	4/47 65.2%
92-102	16/45 43.9%	16/41 37.9%	10/30 30.3%	15/43 42.4%
Terminal 102-104	29	25	20	28

[1] number deaths / number at risk
[2] per cent survival to end of period.

A summary of the results of tests of dose effect on survival in mice are presented below:

Table 22. Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.6054	0.6502	0.1549	0.0948
No Trend over all four groups	0.7949	0.5793	0.9095	0.9852
No difference between high dose and vehicle	0.4564	0.3617	0.7822	0.6885

In male mice, the survival curve for the vehicle eventually slightly dominates the survival curves in the other groups, but this is not sufficient to result in any statistically significant test of homogeneity, dose related trend, or difference between the high dose and control (all 6 $p \geq 0.3617$). In female mice it does seem that the vehicle, low, and high dose groups are eventually

Reference ID: 3401357

intertwined, but with higher survival than the medium dose group. As the Wilcoxon statistic is more sensitive to later differences, this is sufficient to result in test of overall homogeneity that is arguable somewhat close to significance (Wilcoxon p = 0.0948), but less so when weighting earlier difference (Logrank p = 0.1549). However, there is no strong evidence of dose related trend or heterogeneity between the high dose and control (all four p ≥ 0.6885).

Tumorigenicity analysis:

Table 23 below, a repeat of Table 4 above and a synopsis of Table A.2.2 below, shows the tumors that had at least one non-multiplicity adjusted test that was statistically significant at a 0.05 level.

Table 23. Potentially Statistically Significant Neoplasms in Mice

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Mice								
Caecum+Colon+Duodenum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
Adenoma	0	0	0	3	.0134	.1025	.	.
Caecum+Colon+Duodenum+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.6	45.8	46.5	45.4				
Adenoma/Adenocarcinoma	2	0	0	5	.0114	.1764	1	1
DUODENUM								
# Evaluated	66	65	65	66				
Adj. # at Risk	50.1	44.9	46.5	45.1				
ADENOMA	0	0	0	2	.0582	.2217	.	.
Female Mice								
UTERINE CERVIX								
# Evaluated	66	64	65	66				
Adj. # at Risk	48.6	44.3	39.8	48.1				
ENDOMETRIAL POLYP	0	0	1	2	.0652	.2474	.4483	.

In mice the only test that met the multiplicity adjusted limits for statistical significane were adenomas in the caecum or colon or duodenum (i.e the intestine) of male mice (p = 0.0134 < 0.025), while the corresponding pairwise test was close to significance (p = 0.1025 ≈ 0.10). Note, however, if we consider pooled adenoma and adenocarcinoma of those organs plus the jejunum to be rare (ignoring the incidence in the vehicle group) the dose realed trend test, but not the pairwise comparison with control, would also be classified as statistically significant.

Complete results of statistical poly-k tests of tumor trend and differences between dose groups in mice are given in Tables A.2.5 and A.2.6 in Appendix 2.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

Reference ID: 3401357

5. SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Please see Section 1.3 above.

5.2. Conclusions and Recommendations

Please see section 1.1.

APPEARS THIS WAY ON ORIGINAL

Reference ID: 3401357

APPENDICES:

Appendix 1. Survival Analysis

Simple summary life tables in mortality are presented in the report (Tables 14, 15, 20, and 21, above). These curves, defined across study groups for each species by gender combination, are displayed below in Figures A.1.1 through A.1.4. The plots include 95% confidence intervals around each survival curve (colored area around each curve). These plots are also supported by tests of homogeneity in survival over the treatment groups. The statistical significance levels (i.e., p-values) are provided in Tables A.1.1. and A.1.2, below. One might note that the log rank tests place greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus places more weight on differences in earlier events than does the log rank test.

Table A.1.1 Statistical Significances of Tests of Homogeneity and Trend in Survival in Rats

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.0223	0.0082	0.5557	0.3796
No Trend over all four groups	0.1225	0.2021	0.2271	0.2051
No difference between high dose and vehicle	0.4005	0.7710	0.1343	0.3311

The tests of homogeneity in survival over doses in male rats are statistically significant at the usual 0.05 level (Logrank p = 0.0233, Wilcoxon p = 0.0082). From the Kaplan-Meier curve in Figure A.1.1, it seems clear that this is due to the overall lower survival in the low dose group while the medium dose group has generally the highest survival, with survival in the high dose and vehicle crossing, but mostly bounded between these curves. It seems difficult to attribute these results to any particular dose related effect. This seems to be consistent with the lack of any statistically significant dose related trend (Logrank p = 0.1225, Wilcoxon p = 0.2021) or any statistically significant difference between the the high dose and vehicle (Logrank p = 0.4005, Wilcoxon p = 0.7710). In female rats, the Kaplan-Meier curves suggest that through most of the study the vehicle has highest survival, with the other curves rather intertwined. However, these apparent differences were not statistically significant. Whether analyzing homogeneity or trend among all four dose groups or the comparison of the High dose and vehicle, no tests were statistically significance (i.e., all 6 $p \geq 0.1343$). While absence of proof is not proof of absence, the lack of evidence for such differences in survival is consistent with the hypotheses of no dose related differences or trends.

The Kaplan-Meier survival curves in male and female rats rats are presented next.

Reference ID: 3401357

Figure A.1.1 Kaplan-Meier Survival Curves for Male Rats

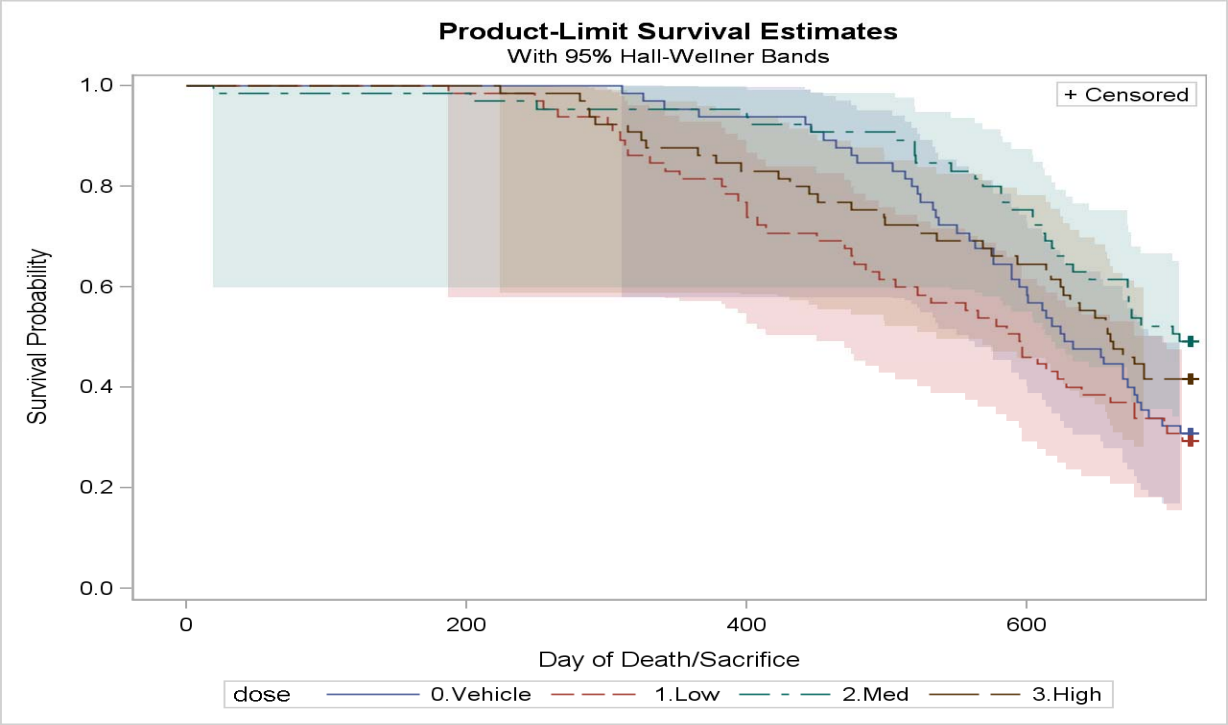
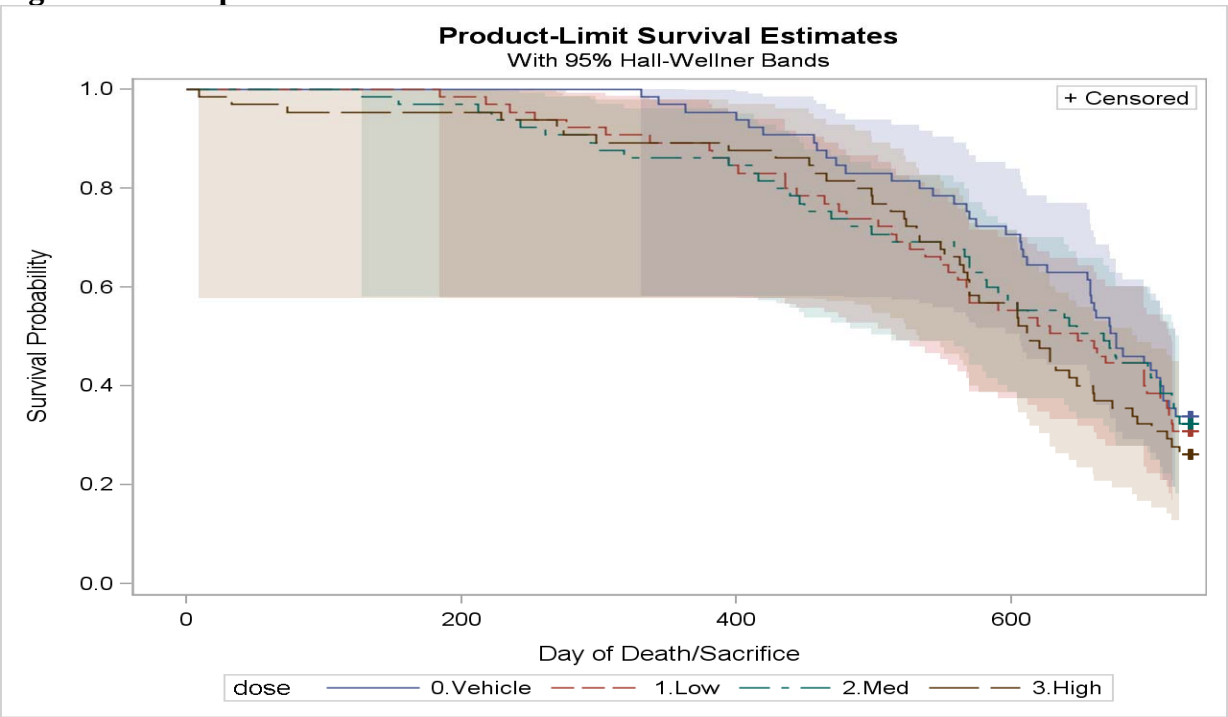


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



Reference ID: 3401357

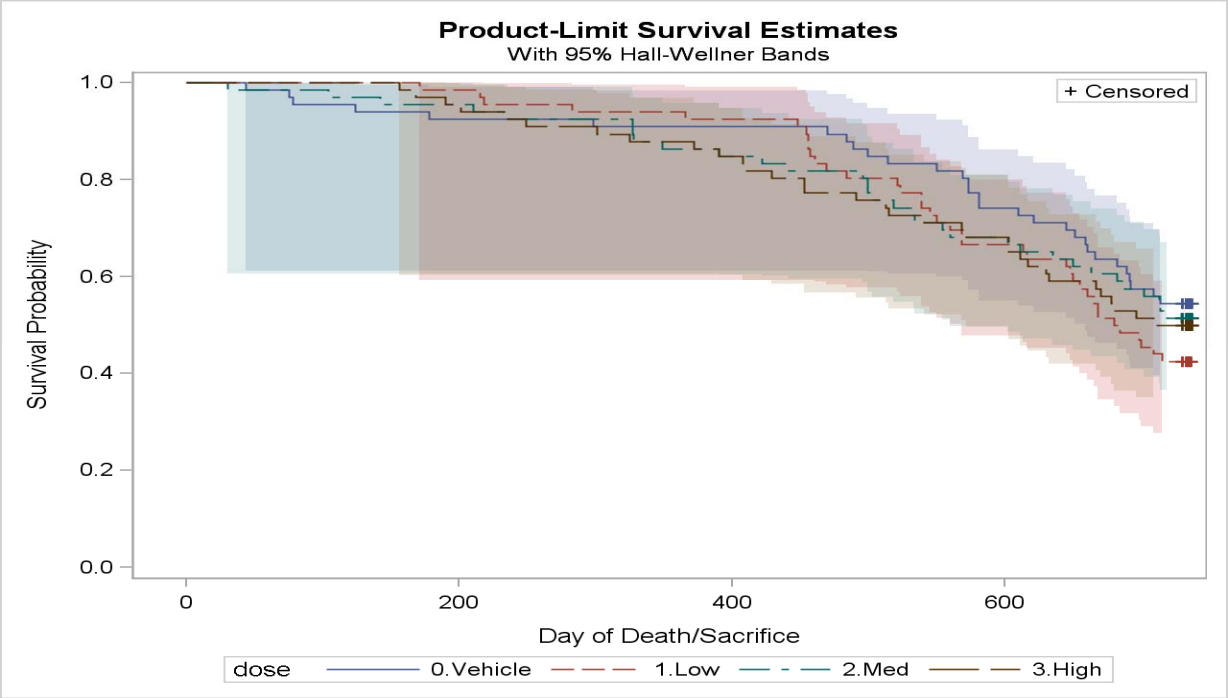
The summary of the results of tests of dose effect on survival in mice are presented below:

Table A.1.2 Statistical Significances of Tests of Homogeneity and Trend in Survival in Mice

Hypotheses	Males		Females	
	Logrank	Wilcoxon	Logrank	Wilcoxon
Homogeneity over all four groups	0.6054	0.6502	0.1549	0.0948
No Trend over all four groups	0.7949	0.5793	0.9095	0.9852
No difference between high dose and vehicle	0.4564	0.3617	0.7822	0.6885

In male mice, the survival curve for the vehicle eventually slightly dominates the survival curves in the other groups, but this is not sufficient to result in any statistically significant test of homogeneity, dose related trend, or difference between the high dose and control (all 6 $p \geq 0.3617$). In female mice it does seem that the vehicle, low, and high dose groups are eventually intertwined, but with higher survival than the medium dose group. As the Wilcoxon statistic is more sensitive to later differences, this is sufficient to result in test of overall homogeneity that is arguable somewhat close to significance (Wilcoxon $p = 0.0948$), but less so when weighting earlier difference (Logrank $p = 0.1549$). However, there is no strong evidence of dose related trend or heterogeneity between the high dose and control (all 4 $p \geq 0.6885$).

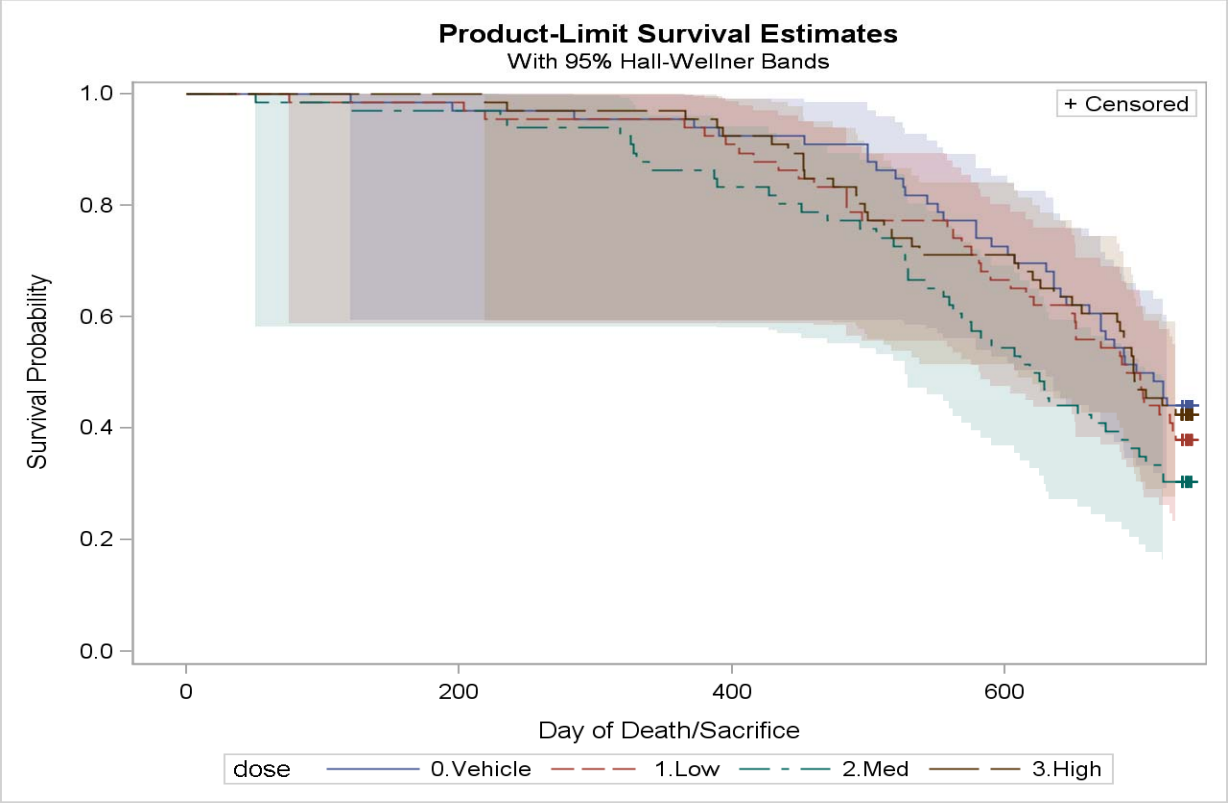
Figure A.1.3 Kaplan-Meier Survival Curves for Male Mice



Reference ID: 3401357

The survival curves in female mice is given below:

Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Reference ID: 3401357

Appendix 2. FDA Poly-k Tumorigenicity Analysis

The poly-k test, here with k=3, modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The tests used here are small sample exact permutation tests of tumor incidence. When there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison an argument could be made that the p-value for this test should be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by a period ".". Note that the StatXact program used for these analyses adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

For each gender by organ the number of animals microscopically analyzed is presented first. Note that indicating an organ was not examined requires a specification in the data (please see section 2.2 above). It is possible that this specification could be missing in some of this data. Then the number of animals at risk could be inflated, and the proportion of animals with tumor would be artificially decreased. Thus, as discussed in Section 1.5 above, for some of these organs it is possibly more appropriate to define the actual endpoint used in the statistical analysis be the condition of being microscopically analyzed AND show the tumor. This does have problems unless treatment groups are not treated equally except for actual treatment. The entry for each tumor is preceded by the adjusted number of animals at risk for that endpoint. It seems clear that an animal that dies early without having displaying that endpoint reduces the size of the risk set for that getting that particular endpoint. The poly-k test down weights such animals, and as discussed in Section 1.3.1.4, above, the sum of these poly-k weights seems to be a better estimate of the number of animals at risk of getting that tumor than the simple number of animals analyzed. This sum is given in the row labeled "Adjusted # at risk ". Tumor incidence is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high and medium dose groups versus vehicle. The next row continues with the p-values of the pairwise test between the low and vehicle dose groups and the p-values between the vehicle dose group and high dose group with water, respectively. For these analyses, incidence in the water only group is used to assess background tumor incidence, and thus whether a tumor is considered to be rare (background incidence <1%) or common. Note that a tumor is only classified as rare if the vehicle group shows none of that particular tumor.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman (HLR) rules discussed in Section 1.3.1.5 are often applied. The question here is whether we want to control type I error over both studies when testing trend and the pairwise difference between the high dose and control jointly or separately. That is, when testing for both trend and pairwise difference together, to control the overall Type I error rate to roughly 10% one compares the unadjusted significance level of the trend test to 0.005 for common tumors and 0.025 for rare tumors, and the pairwise test to 0.05 for common tumors and 0.10 for rare tumors. To compare them separately we use the same bounds for the test of trend, but for the pairwise test use 0.01

Reference ID: 3401357

and 0.05, respectively. Using these adjustments for other tests, like testing the comparisons between the low and medium dose groups versus vehicle can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Table A.2.1 below, shows the tumors that had at least one non-multiplicity adjusted test in rats that was statistically significant at least at a 0.05 level

Table A.2.1. Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.7	36.1	49.9	43.2				
HEPATOCELLULAR ADENOMA	2	2	6	8	.0123	.0444	.1802	.6222
Adj. # at Risk	43.7	36.1	50.7	43.2				
Hepatocellular Adenoma/Carcinoma	2	2	8	8	.0145	.0444	.0746	.6222
PANCREAS								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
ISLET CELL ADENOMA	1	5	5	2	.6541	.5000	.1411	.0692
Adj. # at Risk	42.9	36.1	49.3	43.0				
Islet Cell Adenoma/Carcinoma	2	6	7	2	.7843	.6921	.1212	.0878
PITUITARY								
# Evaluated	63	65	65	64				
Adj. # at Risk	47.6	45.6	55.0	46.9				
ADENOMA, PARS DISTALIS	19	28	30	23	.4822	.2361	.1104	.0296
SKIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.1	35.7	50.3	43.4				
FIBROMA	1	0	5	3	.0920	.3080	.1404	1
Adj. # at Risk	43.1	35.7	50.3	43.9				
Fibroma/Sarcoma NOS/Fibrosarcoma	1	0	5	4	.0407	.1800	.1404	1
TESTES								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.3	35.7	49.2	43.2				
INTERSTITIAL (LEYDIG) CELL ADENOMA	3	1	4	6	.0589	.2417	.5735	.9135
Female Rats								
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.4	39.6	40.5	40.4				
HEPATOCELLULAR ADENOMA	2	1	6	7	.0087	.0539	.0980	.8511
MAMMARY								
# Evaluated	64	65	64	65				
Adj. # at Risk	47.6	42.4	43.5	44.8				
MAMMARY ADENOCARCINOMA	10	10	12	18	.0146	.0356	.3134	.4864
OVARIES								
# Evaluated	64	65	64	63				
Adj. # at Risk	45.2	39.6	39.3	38.2				
SERTOLIFORM TUBULAR ADENOMA	0	0	0	2	.0546	.2066	.	.
Adj. # at Risk	45.2	39.6	39.4	38.2				
Thecal Cell Tmr/Serto.Tub.Adenoma	0	0	1	2	.0525	.2066	.4643	.

Reference ID: 3401357

Table A.2.1. (cont.) Potentially Statistically Significant Neoplasms in Rats

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Female Rats (cont.)								
UTERINE CERVIX								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.2	39.6	39.3	39.3				
SQUAMOUS CELL CARCINOMA	0	0	0	2	.0568	.2126	.	.
UTERUS								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endo. Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722
Adj. # at Risk	45.5	39.6	39.5	39.9				
SQUAMOUS CELL CARCINOMA	1	0	1	3	.0543	.2558	.7160	1
Uterus(w/uterine cervix)								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endometrial Adenoma/-Adenocarcinoma	3	2	2	11	.0005	.0102	.7722	.7722
Addj. # at Risk	45.5	39.6	39.5	40.8				
SQUAMOUS CELL CARCINOMA	1	0	1	5	.0059	.0765	.7160	1

Note that all the tumors in male rats would be classified as common (incidence > 1%). Using the Haseman-Lin-Rahman rules for separate tests A, in male rats no test of trend or pairwise difference between the high dose and control was statistically significant, although the tests for hepatocellular adenoma (trend p = 0.0123 ≈ 0.005, pairwise p = 0.0444 ≈ 0.01) and pooled adenoma and carcinoma (trend p = 0.0145 ≈ 0.005, pairwise p = 0.0444 ≈ 0.01) of the liver were all somewhat close to significance. In female rats, in the table above, sertoliform tubular adenoma, plus pooled thecal cell tumor and sertoliform tubular adenoma, both of the ovaries, and squamous cell carcinoma of the uterine cervix would be classified as rare tumors (but not in the uterus or pooled uterus and cervix), the remainder as common. For tumor incidence combining the uterus with the uterine cervix only affects the results for squamous cell carcinoma. The test of trend in hepatocellular adenoma of the liver would be close to the multiplicity adjusted level of statistical significance (p = 0.0087 ≈ 0.005). The corresponding tests comparing the high dose to vehicle might be classified as close to significance (p = 0.0539 ≈ 0.01). The tests of trend in endometrial adenocarcinoma and pooled adenoma and adenocarcinoma of the uterus (and thus the uterus with uterine cervix) would both be categorized as statistically significant (p = 0.0002, 0.0005 < 0.005, respectively), while the corresponding tests between the high dose and vehicle were statistically significant or close (p = 0.0035 < 0.01 < 0.05 and 0.0102 ≈ 0.01 < 0.05). Strictly following the HLR rules and specifying that squamous cell carcinoma of the uterus with cervix is a common tumor, we would conclude that both the test of trend and the pairwise comparison are close to statistical significance (p = 0.0059 ≈ 0.005 amd 0.0765 ≈ 0.01). No other tests of carcinogenicity between dose groups achieved the multiplicity adjusted levels of significance.

Table A.2.2 below, shows the tumors that had at least one non-multiplicity adjusted test in mice that was statistically significant at a 0.05 level. In mice the only test that met the

Reference ID: 3401357

multiplicity adjusted limits for statistical significane were adenomas in the caecum or colon or duodenum (i.e the intestine) of male mice (p = 0.0134 < 0.025), while the corresponding pairwise test was close to significance (p = 0.1025 ≈ 0.10). Note, however, if we consider pooled adenoma and adenocarcinoma of those organs plus the jejunum to be rare (ignoring the incidence in the vehicle group) the dose realed trend test, but not the pairwise comparison with control, would also be classified as statistically significant.

Table A.2.2. Potentially Statistically Significant Neoplasms in Mice

Sex/ organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
Male Mice								
Caecum+Colon+Duodenum+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.6	45.8	46.5	45.4				
Adenoma/Adenocarcinoma	2	0	0	5	.0114	.1764	1	1
Caecum+Colon+Duodenum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
Adenoma	0	0	0	3	.0134	.1025	.	.
DUODENUM								
# Evaluated	66	65	65	66				
Adj. # at Risk	50.1	44.9	46.5	45.1				
ADENOMA	0	0	0	2	.0582	.2217	.	.
LUNGS + BRONCHI								
# Evaluated	66	66	66	66				
Adj. # at Risk	51.4	50.3	46.8	47.4				
BRONCHIOLOALVEOLAR ADENOMA	19	27	14	18	.7390	.5404	.8218	.0680
Female Mice								
UTERINE CERVIX								
# Evaluated	66	64	65	66				
Adj. # at Risk	48.6	44.3	39.8	48.1				
ENDOMETRIAL POLYP	0	0	1	2	.0652	.2474	.4483	.

Complete Incidence tables are provided below:

Table A.2.3. Tumor Incidence and Results of Tests in Male Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
ABDOMEN								
# Evaluated	2	1	0	1				
Adj. # at Risk	2.0	0.8	0.0	0.6				
MALIGNANT SCHWANNOMA	1	0	0	0	1	.	.	.
Adj. # at Risk	1.5	1.0	0.0	0.6				
MESOTHELIOMA	1	1	0	0	1	.	.	.
ADIPOSE TISSUE								
# Evaluated	7	4	2	1				
Adj. # at Risk	5.2	2.7	1.7	0.8				
FIBROMA	0	0	1	0	.1250	.	.1667	.

Reference ID: 3401357

Table A.2.3 (cont.). Tumor Incidence and Results of Tests in Male Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
ADRENALS								
# Evaluated	65	65	64	65				
Adj. # at Risk	43.3	36.1	49.3	42.8				
CORTICAL ADENOMA	2	2	2	2	.5308	.6831	.7394	.6222
Adj. # at Risk	43.4	35.7	48.7	42.8				
MALIGNANT PHAEOCHROMOCYTOMA	2	0	0	0	1	1	1	1
Adj. # at Risk	44.6	36.0	49.8	43.4				
PHAEOCHROMOCYTOMA	9	6	10	5	.8488	.9219	.6045	.7440
Adj. # at Risk	45.1	36.0	49.8	43.4				
Pheochromocytoma [B&M]	11	6	10	5	.9137	.9681	.7634	.8573
BRAIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.1	36.5	49.9	42.8				
ASTROCYTOMA	1	1	1	0	.8467	1	.7843	.7069
Adj. # at Risk	43.1	37.4	49.9	42.8				
Astrocytoma/Oligodendroglioma	1	2	1	1	.6370	.7471	.7843	.4431
Adj. # at Risk	42.9	35.7	49.2	42.8				
GRANULAR CELL TUMOUR	0	0	0	1	.2500	.5000	.	.
Adj. # at Risk	42.9	36.6	49.2	42.8				
OLIGODENDROGLIOMA	0	1	0	1	.3121	.5000	.	.4615
H-POIETIC TUMOUR								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	36.2	49.6	43.2				
HISTIOCYTIC SARCOMA	0	1	1	1	.3151	.5059	.5385	.4615
Adj. # at Risk	43.7	35.7	49.2	43.8				
MALIGNANT LYMPHOMA	2	0	0	1	.6085	.8794	1	1
Adj. # at Risk	43.1	35.7	49.2	43.7				
MYELOID CELL LEUKAEMIA	1	0	0	1	.4430	.7529	1	1
HEAD								
# Evaluated	2	0	1	2				
Adj. # at Risk	2.0	0.0	0.0	1.1				
SQUAMOUS CELL CARCINOMA, ZYMBAL	2	0	0	1	1	.	.	.
JEJUNUM								
# Evaluated	54	53	59	52				
Adj. # at Risk	36.7	30.8	46.9	36.3				
ADENOMA	0	0	1	0	.5541	.	.5610	.
KIDNEYS								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
TUBULAR CARCINOMA	0	1	0	0	.7500	.	.	.4545
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.7	36.1	49.9	43.2				
HEPATOCELLULAR ADENOMA	2	2	6	8	.0123	.0444	.1802	.6222
Adj. # at Risk	42.9	35.7	50.0	42.8				
HEPATOCELLULAR CARCINOMA	0	0	2	0	.5225	.	.2872	.
Adj. # at Risk	43.7	36.1	50.7	43.2				
Hepatocellular Adenoma/Carcinoma	2	2	8	8	.0145	.0444	.0746	.6222
LN MESENTERIC								
# Evaluated	65	64	65	65				
Adj. # at Risk	42.9	35.5	49.9	42.8				
HAEMANGIOMA	2	0	3	1	.5436	.8795	.5735	1
LUNGS + BRONCHI								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.0	35.7	49.2	42.8				
BRONCHIOLOALVEOLAR ADENOMA	1	0	0	0	1	1	1	1

Reference ID: 3401357

Table A.2.3. (cont.) Tumor Incidence and Results of Tests in Male Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
MAMMARY								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
MAMMARY ADENOCARCINOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	43.8	36.6	49.2	42.8				
MAMMARY FIBROADENOMA	3	1	0	0	.9963	1	1	.9179
Adj. # at Risk	43.0	35.7	49.2	43.1				
SARCOMA NOS	1	0	0	1	.4430	.7529	1	1
PANCREAS								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
ACINAR CELL ADENOMA	0	1	1	2	.1264	.2470	.5385	.4545
Adj. # at Risk	42.9	36.1	49.3	43.0				
ISLET CELL ADENOMA	1	5	5	2	.6541	.5000	.1411	.0692
Adj. # at Risk	42.9	35.7	49.2	42.8				
ISLET CELL CARCINOMA	1	1	2	0	.8142	1	.5582	.7057
Adj. # at Risk	42.9	36.1	49.3	43.0				
Islet Cell Adenoma/Carcinoma	2	6	7	2	.7843	.6921	.1212	.0878
Adj. # at Risk	43.2	35.7	49.2	42.8				
MIXED CELL ADENOMA	1	0	0	0	1	1	1	1
PARATHYROIDS								
# Evaluated	61	60	62	60				
Adj. # at Risk	40.4	33.6	48.3	39.3				
ADENOMA	2	0	2	1	.5661	.8751	.7574	1
PINNAE								
# Evaluated	1	1	3	2				
Adj. # at Risk	1.0	1.0	3.0	1.0				
LEIOMYOMA	0	1	0	0	1	.	.	.
PITUITARY								
# Evaluated	63	65	65	64				
Adj. # at Risk	47.6	45.6	55.0	46.9				
ADENOMA, PARS DISTALIS	19	28	30	23	.4822	.2361	.1104	.0296
Adj. # at Risk	41.7	35.7	49.2	42.3				
ADENOMA, PARS INTERMEDIA	1	1	0	0	.9408	1	1	.7123
PREPUTIAL GLANDS								
# Evaluated	65	65	63	65				
Adj. # at Risk	43.0	35.7	48.4	42.9				
SQUAMOUS CELL PAPILLOMA	1	0	0	1	.4409	.7530	1	1
SALIVARY GLANDS								
# Evaluated	65	62	64	65				
Adj. # at Risk	42.9	35.7	49.0	42.8				
MALIGNANT SCHWANNOMA	0	1	0	0	.7500	.	.	.4545
SKELETAL MUSCLE								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
HAEMANGIOMA	0	0	0	1	.2500	.5000	.	.
Adj. # at Risk	42.9	35.7	49.2	42.8				
OSTEOSARCOMA	0	0	1	0	.5417	.	.5385	.

Reference ID: 3401357

Table A.2.3. (cont.) Tumor Incidence and Results of Tests in Male Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
SKIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	35.7	49.2	42.8				
BASAL CELL CARCINOMA	1	0	1	0	.7914	1	.7897	1
Adj. # at Risk	42.9	35.7	49.2	42.8				
BASAL CELL TUMOUR	0	0	0	1	.2500	.5000	.	.
Adj. # at Risk	42.9	35.7	49.2	42.8				
Basal Cell Tumor/Carcinoma	1	0	1	1	.4223	.7530	.7897	1
Adj. # at Risk	43.1	35.7	50.3	43.4				
FIBROMA	1	0	5	3	.0920	.3080	.1404	1
Adj. # at Risk	43.1	35.7	50.3	43.9				
Fibroma/Sarc.NOS/fibrosarcoma	1	0	5	4	.0407	.1800	.1404	1
Adj. # at Risk	43.7	36.5	49.6	43.4				
KERATOACANTHOMA	8	2	3	5	.6269	.8861	.9858	.9846
Adj. # at Risk	42.9	35.7	49.2	43.3				
SARCOMA NOS	0	0	0	1	.2544	.5059	.	.
Adj. # at Risk	42.9	35.7	49.2	42.8				
SEBACEOUS CELL ADENOMA	0	0	1	0	.5417	.	.5385	.
Adj. # at Risk	43.1	36.8	49.2	42.8				
SQUAMOUS CELL PAPILLOMA	1	3	1	0	.9298	1	.7843	.2435
Adj. # at Risk	43.9	37.6	49.6	43.4				
Sq. Cell Papillpma/Keratocanthoma	9	5	4	5	.8459	.9288	.9808	.8787
Adj. # at Risk	43.1	35.7	49.2	42.8				
TRICHIOEPITHELIOMA	1	0	0	0	1	1	1	1
SPINAL C. CERV.								
# Evaluated	65	65	65	65				
Adj. # at Risk	42.9	36.6	49.2	42.8				
ASTROCYTOMA	0	1	0	0	.7515	.	.	.4615
STOMACH								
# Evaluated	64	64	65	65				
Adj. # at Risk	42.6	35.6	49.2	43.1				
SQUAMOUS CELL CARCINOMA	0	0	0	1	.2544	.5059	.	.
Systemic								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.4	36.4	49.2	42.8				
Schwannoma [M]	1	1	0	0	.9371	1	1	.7069
TAIL								
# Evaluated	3	2	2	5				
Adj. # at Risk	2.0	2.0	2.0	4.9				
FIBROSARCOMA	0	0	0	1	.5000	.6667	.	.
TESTES								
# Evaluated	65	65	65	65				
Adj. # at Risk	43.3	35.7	49.2	43.2				
INTERSTITIAL (LEYDIG) CELL ADENOMA	3	1	4	6	.0589	.2417	.5735	.9135
THYMUS								
# Evaluated	60	62	63	62				
Adj. # at Risk	40.1	33.2	47.5	41.0				
THYMOMA (EPITHELIAL)	1	0	0	0	1	1	1	1

Reference ID: 3401357

Table A.2.3. (cont.) Tumor Incidence and Results of Tests in Male Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
THYROIDS								
# Evaluated	65	64	64	65				
Adj. # at Risk	43.9	35.9	49.0	43.7				
C-CELL ADENOMA	5	2	1	4	.5090	.7583	.9915	.9070
Adj. # at Risk	42.9	35.7	49.0	42.8				
C-CELL CARCINOMA	0	0	0	1	.2500	.5000	.	.
Adj. # at Risk	43.9	35.9	49.0	43.7				
C-cell Adenoma/Carcinoma.	5	2	1	5	.3379	.6308	.9915	.9070
Adj. # at Risk	42.9	35.7	49.0	42.8				
FOLLICULAR CELL ADENOMA	0	0	1	0	.5417	.	.5385	.
Adj. # at Risk	42.9	35.7	49.0	42.8				
FOLLICULAR CELL CARCINOMA	0	0	0	1	.2500	.5000	.	.
Adj. # at Risk	42.9	35.7	49.0	42.8				
Foll. Cell Adenoma/Carcinoma.	0	0	1	1	.2081	.5000	.5385	.

Table A.2.4. Tumor Incidence and Results of Tests in Female Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
ABDOMEN								
# Evaluated	0	1	2	4				
Adj. # at Risk	0.0	1.0	1.9	2.9				
MESOTHELIOMA	0	1	0	0	1	.	.	.
Adj. # at Risk	0.0	0.5	2.0	2.9				
SARCOMA NOS	0	0	1	0	1	.	.	.
ADRENALS								
# Evaluated	65	65	65	64				
Adj. # at Risk	45.6	39.7	40.3	37.5				
CORTICAL ADENOMA	1	2	1	0	.8501	1	.7227	.4459
Adj. # at Risk	46.2	39.6	40.2	37.5				
MALIGNANT PHAEOCHROMOCYTOMA	2	0	0	0	1	1	1	1
Adj. # at Risk	47.6	39.6	40.9	38.6				
PHAEOCHROMOCYTOMA	11	1	3	3	.9032	.9888	.9914	.9997
Adj. # at Risk	48.5	39.6	40.9	38.6				
Pheochromocytoma [B&M]	13	1	3	3	.9530	.9960	.9971	.9999
BRAIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.6	40.2	38.5				
ASTROCYTOMA	0	0	0	1	.2346	.4578	.	.
FEMUR INC. JOINT								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.6	40.3	38.4				
OSTEOMA	0	0	1	0	.4815	.	.4706	.
H-POIETIC TUMOUR								
# Evaluated	65	65	65	65				
Adj. # at Risk	46.2	39.6	40.2	38.5				
MALIGNANT LYMPHOMA	1	0	0	1	.4130	.7031	1	1
Adj. # at Risk	45.3	39.6	40.4	38.4				
MYELOID CELL LEUKAEMIA	0	0	1	0	.4815	.	.4706	.
HEAD								
# Evaluated	1	1	0	0				
Adj. # at Risk	0.9	1.0	0.0	0.0				
SQUAMOUS CELL CARCINOMA, ZYMBA	0	1	0	0	1	.	.	.
Adj. # at Risk	1.0	0.3	0.0	0.0				
SQUAMOUS CELL CARCINOMA-ORAL C	1	0	0	0	1	.	.	.

Reference ID: 3401357

Table A.2.4. (cont.) Tumor Incidence and Results of Tests in Female Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
HEART								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.6	40.2	38.7				
MALIGNANT SCHWANNOMA	0	0	0	1	.2346	.4578	.	.
JEJUNUM								
# Evaluated	61	63	61	63				
Adj. # at Risk	44.4	39.5	38.8	38.2				
LEIOMYOSARCOMA	0	0	1	0	.4780	.	.4634	.
LIVER								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.4	39.6	40.5	40.4				
HEPATOCELLULAR ADENOMA	2	1	6	7	.0087	.0539	.0980	.8511
LN MESENTERIC								
# Evaluated	65	65	65	63				
Adj. # at Risk	45.4	39.6	40.2	37.9				
HAEMANGIOMA	1	1	0	0	.9231	1	1	.7160
MAMMARY								
# Evaluated	64	65	64	65				
Adj. # at Risk	52.2	50.7	46.5	48.3				
Adenoma/Adenocarc./Fibroad.	33	35	26	33	.4178	.3649	.8178	.3123
Adj. # at Risk	47.6	42.4	43.5	44.8				
MAMMARY ADENOCARCINOMA	10	10	12	18	.0146	.0356	.3134	.4864
Adj. # at Risk	45.6	39.7	39.6	39.2				
MAMMARY ADENOMA	2	2	2	4	.1372	.2721	.6360	.6360
Adj. # at Risk	49.9	48.5	43.0	44.0				
MAMMARY FIBROADENOMA	25	30	17	22	.7415	.5774	.9056	.1747
MESENTRY								
# Evaluated	0	0	0	1				
Adj. # at Risk	0.0	0.0	0.0	1.0				
LEIOMYOSARCOMA	0	0	0	1	1	.	.	.
Mesentry+Jejunum								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.6	40.3	39.0				
Leiomyosarcoma	0	0	1	1	.1743	.4643	.4706	.
OVARIES								
# Evaluated	64	65	64	63				
Adj. # at Risk	45.2	39.6	39.3	38.2				
SERTOLIFORM TUBULAR ADENOMA	0	0	0	2	.0546	.2066	.	.
Adj. # at Risk	45.2	39.6	39.4	37.9				
THECAL CELL TUMOUR	0	0	1	0	.4750	.	.4643	.
Adj. # at Risk	45.2	39.7	39.3	37.9				
TUBULAR ADENOMA	0	1	0	0	.7188	.	.	.4643
Adj. # at Risk	45.2	39.6	39.4	38.2				
Thecal Cell Tmr/Serto.Tub.Adenoma	0	0	1	2	.0525	.2066	.4643	.
PANCREAS								
# Evaluated	65	65	65	64				
Adj. # at Risk	45.6	39.8	40.2	38.4				
ISLET CELL ADENOMA	2	2	0	0	.9746	1	1	.6360
PARATHYROIDS								
# Evaluated	61	63	62	60				
Adj. # at Risk	43.6	38.6	39.0	34.6				
ADENOMA	1	0	0	0	1	1	1	1

Reference ID: 3401357

Table A.2.4. (cont.) Tumor Incidence and Results of Tests in Female Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
PITUITARY								
# Evaluated	65	65	65	64				
Adj. # at Risk	60.9	57.7	58.6	53.1				
ADENOMA, PARS DISTALIS	50	52	50	46	.5063	.4032	.4300	.1588
Adj. # at Risk	45.3	39.6	40.2	38.1				
CARCINOMA, PARS DISTALIS	0	0	0	1	.2346	.4578	.	.
Adj. # at Risk	60.9	57.7	58.6	53.2				
Pars Dist. Adenoma/Carcinoma	50	52	50	47	.3764	.2951	.4300	.1588
SKELETAL MUSCLE								
# Evaluated	64	65	65	64				
Adj. # at Risk	45.3	39.6	40.2	38.4				
SARCOMA NOS	1	0	0	0	1	1	1	1
SKIN								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.6	40.3	39.0				
BASAL CELL TUMOUR	0	0	1	1	.1705	.4578	.4706	.
Adj. # at Risk	45.3	39.6	40.2	38.6				
FIBROMA	0	0	0	1	.2346	.4578	.	.
Adj. # at Risk	45.3	39.6	40.3	38.4				
KERATOACANTHOMA	0	0	1	0	.4815	.	.4706	.
Adj. # at Risk	45.3	39.8	40.2	38.4				
SEBACEOUS CELL ADENOMA	0	1	0	0	.7222	.	.	.4643
Adj. # at Risk	45.3	39.6	40.3	38.8				
SQUAMOUS CELL PAPILLOMA	0	0	1	1	.1705	.4578	.4706	.
Adj. # at Risk	45.3	39.6	40.3	38.8				
Sq. Cell Papilloma/Keratocanth	0	0	2	1	.1801	.4578	.2185	.
Adj. # at Risk	45.4	39.6	40.2	38.4				
TRICHIOEPITHELIOMA	1	0	0	0	1	1	1	1
SPLEEN								
# Evaluated	65	65	65	64				
Adj. # at Risk	45.6	39.6	40.2	38.4				
HAEMANGIOSARCOMA	1	0	0	0	1	1	1	1
Systemic								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.7	39.6	40.2	38.4				
Hemangioma/Hemangiosarcoma	2	1	0	0	.9796	1	1	.8511
Adj. # at Risk	45.3	39.6	41.1	38.4				
Schwannoma [M]	0	0	1	0	.4847	.	.4767	.
THYMUS								
# Evaluated	64	61	64	63				
Adj. # at Risk	44.4	36.4	39.6	37.6				
THYMOMA (LYMPHOID)	0	1	1	0	.5967	.	.4699	.4500
THYROIDS								
# Evaluated	65	65	65	65				
Adj. # at Risk	46.3	41.5	40.3	39.2				
C-CELL ADENOMA	4	4	2	2	.8000	.8563	.8635	.5765
Adj. # at Risk	45.4	39.6	40.2	38.5				
C-CELL CARCINOMA	1	0	0	1	.4152	.7091	1	1
Adj. # at Risk	46.4	41.5	40.3	39.2				
C-cell Adenoma/Carcinoma.	5	4	2	3	.7122	.8070	.9203	.6968
Adj. # at Risk	45.4	39.6	40.2	38.5				
FOLLICULAR CELL ADENOMA	1	0	0	1	.4152	.7091	1	1
TONGUE								
# Evaluated	65	65	65	65				
Adj. # at Risk	45.3	39.7	40.2	38.4				
SQUAMOUS CELL PAPILLOMA	0	1	0	0	.7222	.	.	.4643

Reference ID: 3401357

Table A.2.4. (cont.) Tumor Incidence and Results of Tests in Female Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
URINARY BLADDER								
# Evaluated	64	65	65	64				
Adj. # at Risk	45.8	39.6	40.2	38.4				
TRANSITIONAL CELL PAPILLOMA	1	0	0	0	1	1	1	1
UTERINE CERVIX								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.2	39.6	39.8	38.6				
GRANULAR CELL TUMOUR	0	1	1	1	.2649	.4578	.4643	.4643
Adj. # at Risk	45.7	39.6	39.3	38.4				
POLYP	1	0	0	0	1	1	1	1
Adj. # at Risk	45.7	39.6	39.3	39.3				
Polyp/Squamous Cell Carcinoma	1	0	0	2	.1442	.4459	1	1
Adj. # at Risk	45.2	39.6	39.3	39.3				
SQUAMOUS CELL CARCINOMA	0	0	0	2	.0568	.2126	.	.
UTERUS								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.2	39.6	39.3	38.4				
ENDOMETRIAL ADENOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	46.5	39.7	39.3	39.6				
ENDOMETRIAL POLYP	5	4	1	5	.3860	.5206	.9786	.6684
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endo. Adenoma/-Adenocarc.	3	2	2	11	.0005	.0102	.7722	.7722
Adj. # at Risk	45.2	39.6	39.3	38.5				
LEIOMYOMA	0	0	0	1	.2360	.4578	.	.
Adj. # at Risk	45.2	39.6	40.1	38.4				
MALIGNANT SCHWANNOMA	0	0	1	0	.4815	.	.4706	.
Adj. # at Risk	45.5	39.6	39.5	39.9				
SQUAMOUS CELL CARCINOMA	1	0	1	3	.0543	.2558	.7160	1
Uterus(w/cervix)								
# Evaluated	64	65	64	64				
Adj. # at Risk	45.3	39.7	39.3	40.6				
ENDOMETRIAL ADENOCARCINOMA	2	2	2	11	.0002	.0035	.6360	.6360
Adj. # at Risk	45.2	39.6	39.3	38.4				
ENDOMETRIAL ADENOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	46.5	39.7	39.3	39.6				
ENDOMETRIAL POLYP	5	4	1	5	.3860	.5206	.9786	.6684
Adj. # at Risk	45.3	39.7	39.3	40.6				
Endo. Adenoma/-Adenocarc.	3	2	2	11	.0005	.0102	.7722	.7722
Adj. # at Risk	45.2	39.6	39.8	38.6				
GRANULAR CELL TUMOUR	0	1	1	1	.2649	.4578	.4643	.4643
Adj. # at Risk	45.2	39.6	39.3	38.5				
LEIOMYOMA	0	0	0	1	.2360	.4578	.	.
Adj. # at Risk	45.2	39.6	40.1	38.4				
MALIGNANT SCHWANNOMA	0	0	1	0	.4815	.	.4706	.
Adj. # at Risk	45.7	39.6	39.3	38.4				
POLYP	1	0	0	0	1	1	1	1
Adj. # at Risk	45.7	39.6	39.3	39.3				
Polyp/Squamous Cell Carcinoma	1	0	0	2	.1442	.4459	1	1
Adj. # at Risk	45.5	39.6	39.5	40.8				
SQUAMOUS CELL CARCINOMA	1	0	1	5	.0059	.0765	.7160	1

Reference ID: 3401357

Table A.2.4. (cont.) Tumor Incidence and Results of Tests in Female Rats

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
VAGINA								
# Evaluated	63	65	63	63				
Adj. # at Risk	44.4	40.1	38.3	36.9				
GRANULAR CELL TUMOUR	0	1	0	0	.7215	.	.	.4762
Adj. # at Risk	44.4	39.6	38.3	36.9				
POLYP	0	0	0	1	.2293	.4500	.	.

Table A.2.5. Tumor Incidence and Results of Tests in Male Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
ADIPOSE TISSUE								
# Evaluated	4	6	4	0				
Adj. # at Risk	3.9	4.8	3.2	0.0				
HAEMANGIOMA	1	1	0	0	.9333	.	1	.8571
Adj. # at Risk	3.9	4.3	3.2	0.0				
HIBERNOMA	0	1	0	0	.7000	.	.	.5714
Adj. # at Risk	3.9	4.2	3.3	0.0				
OSTEOSARCOMA	0	0	1	0	.3000	.	.5000	.
ADRENALS								
# Evaluated	66	66	66	64				
Adj. # at Risk	50.1	45.8	46.5	44.6				
CORTICAL ADENOMA	1	1	0	1	.5076	.7197	1	.7256
Adj. # at Risk	50.1	46.1	47.2	44.4				
SUBCAPSULAR CELL ADENOMA	4	2	2	0	.9752	1	.8831	.8777
BONE								
# Evaluated	13	10	6	7				
Adj. # at Risk	12.5	7.5	5.4	4.7				
OSTEOMA	0	1	0	0	.5714	.	.	.3684
CAECUM								
# Evaluated	66	66	64	65				
Adj. # at Risk	50.6	45.8	46.0	45.1				
ADENOCARCINOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	50.1	45.8	46.0	45.1				
ADENOMA	0	0	0	1	.2432	.4737	.	.
Adj. # at Risk	50.6	45.8	46.0	45.1				
Adenoma/Adenocarcinoma	1	0	0	1	.4283	.7256	1	1
COLON								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.4				
ADENOCARCINOMA	1	0	0	1	.4263	.7256	1	1
Adj. # at Risk	50.1	45.8	46.5	45.1				
ADENOMA	0	0	0	1	.2419	.4737	.	.
Caecum+Colon+Duodenum+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.6	45.8	46.5	45.4				
Adenoma/Adenocarcinoma	2	0	0	5	.0114	.1764	1	1
Caecum+Colon+Duodenum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
Adenoma	0	0	0	3	.0134	.1025	.	.
Caecum+Colon+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.4				
Adenocarcinoma	1	0	0	2	.1457	.4601	1	1

Reference ID: 3401357

Table A.2.5. (cont.) Tumor Incidence and Results of Tests in Male Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
DUODENUM								
# Evaluated	66	65	65	66				
Adj. # at Risk	50.1	44.9	46.5	45.1				
ADENOMA	0	0	0	2	.0582	.2217	.	.
EPIDIDYMIDES								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.6	45.1				
INTERSTITIAL (LEYDIG) ADENOMA	0	0	1	0	.4892	.	.4792	.
GALL BLADDER								
# Evaluated	59	58	61	59				
Adj. # at Risk	45.6	42.5	46.0	43.4				
PAPILLOMA	0	2	2	0	.7390	.	.2527	.2302
H-POIETIC TUMOUR								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.4				
HISTIOCYTIC SARCOMA	0	1	0	1	.2956	.4737	.	.4737
Adj. # at Risk	51.2	47.4	48.0	46.4				
LYMPHOMA	5	4	5	3	.7070	.8298	.5902	.7140
Adj. # at Risk	50.1	47.2	47.4	45.8				
MYELOID CELL LEUKAEMIA	0	3	1	1	.5208	.4737	.4845	.1100
HARDERIAN GLANDS								
# Evaluated	66	66	65	66				
Adj. # at Risk	51.9	47.3	46.9	48.5				
ADENOMA	12	7	14	7	.7911	.9177	.2954	.9100
JEJUNUM								
# Evaluated	66	65	65	66				
Adj. # at Risk	50.1	44.9	46.2	45.1				
ADENOCARCINOMA	0	0	0	1	.2432	.4737	.	.
KIDNEYS								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	46.4	46.5	45.1				
SCHWANNOMA	0	1	0	0	.7326	.	.	.4792
Adj. # at Risk	50.1	45.8	46.5	45.1				
TUBULAR ADENOMA	0	1	1	1	.2781	.4737	.4792	.4737
LACHRYMAL GLDS								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.3				
ADENOMA	0	0	0	1	.2419	.4737	.	.
LIVER								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
HAEMANGIOMA	0	3	0	0	.8823	.	.	.1025
Adj. # at Risk	50.1	45.8	46.5	45.1				
HEPATOBLASTOMA	0	0	1	0	.4892	.	.4792	.
Adj. # at Risk	50.2	46.2	46.6	45.3				
HEPATOCELLULAR ADENOMA	4	3	4	5	.2393	.4326	.5947	.7463
Adj. # at Risk	50.1	45.9	46.5	45.1				
HEPATOCELLULAR CARCINOMA	0	1	0	1	.2956	.4737	.	.4737
Adj. # at Risk	50.2	46.4	46.6	45.3				
Hepatocellular Adenoma/Carcinoma	4	4	4	6	.1793	.3045	.5947	.5947
LN MESENTERIC								
# Evaluated	66	65	66	65				
Adj. # at Risk	50.1	45.1	46.5	44.8				
HAEMANGIOMA	0	0	1	0	.4865	.	.4792	.

Reference ID: 3401357

Table A.2.5. (cont.) Tumor Incidence and Results of Tests in Male Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
LUNGS + BRONCHI								
# Evaluated	66	66	66	66				
Adj. # at Risk	51.4	50.3	46.8	47.4				
BRONCHIOLOALVEOLAR ADENOMA	19	27	14	18	.7390	.5404	.8218	.0680
Adj. # at Risk	50.9	47.1	48.0	45.4				
BRONCHIOLOALVEOLAR CARCINOMA	8	5	11	5	.6675	.8390	.2710	.8584
Adj. # at Risk	52.3	51.7	48.3	47.7				
Bronchioalv. Adenoma/Carcinoma	26	31	23	22	.8040	.6978	.6584	.1834
PANCREAS								
# Evaluated	66	66	65	66				
Adj. # at Risk	50.1	45.8	45.6	45.1				
ISLET CELL ADENOMA	1	0	0	0	1	1	1	1
PARATHYROIDS								
# Evaluated	56	52	48	51				
Adj. # at Risk	41.6	37.4	35.1	37.0				
CHIEF CELL ADENOMA	0	1	0	0	.7267	.	.	.4744
PAWS								
# Evaluated	3	0	0	2				
Adj. # at Risk	1.9	0.0	0.0	2.0				
SQUAMOUS CELL PAPILLOMA	0	0	0	2	.3333	.3333	.	.
PINNAE								
# Evaluated	3	3	3	3				
Adj. # at Risk	2.5	2.2	0.9	1.2				
SQUAMOUS CELL CARCINOMA	0	1	0	0	.6000	.	.	.5000
PREPUTIAL GLANDS								
# Evaluated	66	63	66	66				
Adj. # at Risk	50.1	44.1	46.5	45.2				
HAEMANGIOMA	0	0	0	1	.2432	.4737	.	.
PROSTATE								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
ADENOMA	1	0	0	0	1	1	1	1
Pinnae+Skin+Paws								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.3	45.8	46.5	46.4				
Squamous Cell Carc./-Pap./Kerato.	1	1	0	2	.2281	.4684	1	.7256
SALIVARY GLANDS								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
ADENOMA	1	0	0	0	1	1	1	1
SEMINAL VESICLES								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
SARCOMA NOS	0	0	1	0	.4892	.	.4792	.
SKIN								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.7	45.8	47.1	45.1				
FIBROSARCOMA	1	0	2	0	.6753	1	.4766	1
Adj. # at Risk	50.3	45.8	46.5	45.1				
KERATOACANTHOMA	1	0	0	0	1	1	1	1
SPLEEN								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
HAEMANGIOMA	0	0	1	1	.1779	.4737	.4792	.
STOMACH								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	46.5	45.1				
ADENOMA	0	0	1	0	.4892	.	.4792	.

Reference ID: 3401357

Table A.2.5. (cont.) Tumor Incidence and Results of Tests in Male Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
TESTES								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	45.8	47.2	45.1				
HAEMANGIOMA	0	0	1	0	.4920	.	.4845	.
Adj. # at Risk	50.1	45.8	46.7	45.1				
INTERSTITIAL (LEYDIG) CELL ADENOMA	1	1	3	2	.2695	.4601	.2776	.7256
THORAX								
# Evaluated	0	2	1	3				
Adj. # at Risk	0.0	0.9	0.3	1.4				
MESOTHELIOMA	0	0	0	1	1	.	.	.
Adj. # at Risk	0.0	1.7	0.3	0.5				
OSTEOSARCOMA	0	1	0	0	1	.	.	.
THYROIDS								
# Evaluated	65	66	66	66				
Adj. # at Risk	49.1	46.4	46.5	45.1				
FOLLICULAR CELL ADENOMA	0	1	0	1	.2982	.4787	.	.4842

Table A.2.6. Tumor Incidence and Results of Tests in Female Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
ADRENALS								
# Evaluated	65	66	66	65				
Adj. # at Risk	47.7	45.7	39.3	46.0				
PHAEOCHROMOCYTOMA	0	0	0	1	.2599	.4946	.	.
Adj. # at Risk	47.7	45.7	39.3	46.0				
SUBCAPSULAR CELL ADENOMA	0	0	0	1	.2599	.4946	.	.
Adj. # at Risk	47.7	45.7	39.3	46.0				
SUBCAPSULAR CELL CARCINOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	47.7	45.7	39.3	46.0				
Subcaps. Cell Adenoma/Carcinom	1	0	0	1	.4533	.7473	1	1
BONE								
# Evaluated	7	4	5	0				
Adj. # at Risk	5.5	4.0	3.5	0.0				
LEIOMYOMA	0	0	1	0	.2727	.	.3750	.
CAECUM								
# Evaluated	65	66	64	65				
Adj. # at Risk	47.9	45.7	37.7	46.7				
LEIOMYOMA	0	1	0	0	.7314	.	.	.4891
CLITORAL GLANDS								
# Evaluated	63	63	61	63				
Adj. # at Risk	47.3	44.0	38.3	44.2				
CARCINOMA	0	0	1	0	.4740	.	.4471	.
Caecum+Jejunum								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	46.6	39.3	47.0				
Leiomyoma/Leiomyosarcoma	0	2	0	0	.7906	.	.	.2368
GALL BLADDER								
# Evaluated	63	61	59	57				
Adj. # at Risk	46.8	43.0	35.8	42.5				
PAPILLOMA	1	0	1	0	.7141	1	.6806	1
GENERAL COMMENTS								
# Evaluated	0	1	1	0				
Adj. # at Risk	0.0	1.0	0.4	0.0				
CARCINOMA	0	1	0	0	1	.	.	.

Reference ID: 3401357

Table A.2.6. (cont.) Tumor Incidence and Results of Tests in Female Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
H-POIETIC TUMOUR								
# Evaluated	66	66	66	66				
Adj. # at Risk	50.1	47.1	40.0	48.5				
HISTIOCYTIC SARCOMA	5	4	1	5	.4009	.6033	.9727	.7247
Adj. # at Risk	51.7	51.6	43.1	49.7				
LYMPHOMA	12	14	15	10	.7448	.7312	.1627	.4103
Adj. # at Risk	49.2	46.5	39.3	47.0				
MYELOID CELL LEUKAEMIA	1	1	0	0	.9270	1	1	.7366
HARDERIAN GLANDS								
# Evaluated	65	66	66	66				
Adj. # at Risk	48.4	45.7	39.3	47.0				
ADENOCARCINOMA	0	1	0	0	.7303	.	.	.4839
Adj. # at Risk	49.5	47.6	39.5	47.2				
ADENOMA	6	7	3	4	.8010	.8240	.8542	.4674
Adj. # at Risk	49.5	47.7	39.5	47.2				
Adenoma/Adenocarcinoma	6	8	3	4	.8364	.8240	.8542	.3545
JEJUNUM								
# Evaluated	64	66	64	65				
Adj. # at Risk	47.2	45.7	38.0	47.0				
ADENOMA	0	0	0	1	.2614	.4946	.	.
Adj. # at Risk	47.2	46.6	38.0	46.8				
LEIOMYOSARCOMA	0	1	0	0	.7345	.	.	.4946
LIVER								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.7	39.3	47.0				
HAEMANGIOMA	0	1	0	0	.7303	.	.	.4839
Adj. # at Risk	48.6	46.0	39.3	47.0				
HAEMANGIOSARCOMA	0	1	0	0	.7318	.	.	.4894
Adj. # at Risk	48.6	45.8	39.4	47.0				
HEPATOCELLULAR ADENOMA	0	1	1	0	.6096	.	.4483	.4839
LUNGS + BRONCHI								
# Evaluated	66	65	66	66				
Adj. # at Risk	51.3	45.9	41.5	50.4				
BRONCHIOLOALVEOLAR ADENOMA	16	9	9	14	.4313	.7217	.8941	.9340
Adj. # at Risk	49.0	45.2	41.6	47.5				
BRONCHIOLOALVEOLAR CARCINOMA	4	4	5	4	.5123	.6309	.3994	.6059
Adj. # at Risk	51.7	46.1	43.8	50.9				
Bronchioalveolar Adenoma/Carcinoma	20	13	14	18	.4408	.7048	.8116	.9121
MAMMARY								
# Evaluated	66	66	66	66				
Adj. # at Risk	49.9	46.3	40.8	47.0				
Adenocanthoma/Adenocarcinoma	3	3	6	1	.8274	.9334	.1521	.6307
Adj. # at Risk	49.5	46.3	39.5	47.0				
MAMMARY ADENOACANTHOMA	1	1	3	1	.5167	.7366	.2273	.7366
Adj. # at Risk	48.9	45.8	40.6	47.0				
MAMMARY ADENOCARCINOMA	2	2	3	0	.9092	1	.4132	.6668
Adj. # at Risk	48.9	45.7	39.3	47.0				
MAMMARY FIBROADENOMA	1	0	0	0	1	1	1	1

Reference ID: 3401357

Table A.2.6. (cont.) Tumor Incidence and Results of Tests in Female Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
OVARIES								
# Evaluated	65	66	66	66				
Adj. # at Risk	47.9	45.7	39.3	47.0				
CYSTADENOMA	1	0	1	0	.7313	1	.7042	1
Adj. # at Risk	47.9	45.7	39.3	47.0				
Cystadenoma/Tubular Adenoma	2	0	1	0	.8618	1	.8416	1
Adj. # at Risk	47.6	45.8	39.3	47.0				
GRANULOSA CELL TUMOUR	0	2	0	0	.7948	.	.	.2365
Adj. # at Risk	47.7	45.7	39.3	47.0				
HAEMANGIOMA	1	0	1	0	.7313	1	.7042	1
Adj. # at Risk	47.6	45.7	39.3	47.0				
LEIOMYOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	48.7	45.7	39.3	47.2				
LUTEOMA	7	0	0	3	.6659	.9512	1	1
Adj. # at Risk	48.7	45.8	39.3	47.2				
Luteoma/Granulosa Cell Tumor	7	2	0	3	.7958	.9512	1	.9806
Adj. # at Risk	47.6	45.7	39.3	47.0				
TUBULAR ADENOMA	1	0	0	0	1	1	1	1
PANCREAS								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.8	39.3	47.0				
HAEMANGIOMA	0	1	0	0	.7303	.	.	.4839
Adj. # at Risk	48.6	46.0	39.4	47.0				
ISLET CELL ADENOMA	2	1	1	0	.9111	1	.8368	.8667
PITUITARY								
# Evaluated	66	66	66	65				
Adj. # at Risk	48.6	45.8	39.4	46.8				
ADENOMA, PARS DISTALIS	1	3	1	2	.4437	.4839	.6985	.2843
SALIVARY GLANDS								
# Evaluated	66	64	65	65				
Adj. # at Risk	48.6	44.5	38.9	47.5				
ADENOCARCINOMA	0	0	0	1	.2655	.4947	.	.
SKELETAL MUSCLE								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.9	39.3	47.0				
FIBROSARCOMA	0	1	0	0	.7303	.	.	.4839
Adj. # at Risk	48.6	46.6	39.9	47.3				
SARCOMA NOS	0	1	1	2	.1200	.2421	.4483	.4894
Adj. # at Risk	48.6	46.8	39.9	47.3				
Sarcoma NOS/Fibrosarcoma	0	2	1	2	.1991	.2421	.4483	.2368
SKIN								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.8	45.7	39.3	47.0				
BASAL CELL CARCINOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	49.2	45.7	39.7	47.0				
FIBROSARCOMA	1	0	1	0	.7256	1	.6928	1
Adj. # at Risk	49.3	45.7	39.3	47.0				
HAEMANGIOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	48.6	45.9	39.3	47.0				
KERATOACANTHOMA	0	1	0	0	.7303	.	.	.4839
Adj. # at Risk	48.6	45.7	39.3	47.3				
SARCOMA NOS	0	0	0	1	.2626	.4947	.	.
Adj. # at Risk	49.2	45.7	39.7	47.3				
Sarcoma NOS/Fibrosarcoma	1	0	1	1	.3841	.7421	.6928	1

Reference ID: 3401357

Table A.2.6. (cont.) Tumor Incidence and Results of Tests in Female Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
SPLEEN								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	46.2	39.3	47.0				
HAEMANGIOMA	0	2	0	0	.7906	.	.	.2368
Adj. # at Risk	48.6	45.7	39.3	47.0				
HAEMANGIOSARCOMA	0	1	0	0	.7303	.	.	.4839
STOMACH								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.7	39.3	47.8				
SQUAMOUS CELL CARCINOMA	0	0	0	1	.2626	.4947	.	.
Adj. # at Risk	48.6	45.7	39.8	47.0				
SQUAMOUS CELL PAPILLOMA	0	0	1	0	.4775	.	.4483	.
Adj. # at Risk	48.6	45.7	39.8	47.8				
Sq.Cell Papilloma/Carcinoma	0	0	1	1	.1829	.4947	.4483	.
Skin + Tail								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.7	39.3	47.0				
Sarcoma NOS/Fibrosarcoma	0	0	0	1	.2584	.4894	.	.
Systemic								
# Evaluated	66	66	66	66				
Adj. # at Risk	49.2	46.2	39.4	47.1				
HAEMANGIOMA	4	4	2	1	.9361	.9688	.8370	.6066
Adj. # at Risk	49.8	46.0	39.3	47.4				
HAEMANGIOSARCOMA	2	2	0	2	.4840	.6756	1	.6672
Adj. # at Risk	50.4	46.5	39.4	47.6				
Hemangioma/Hemangiosarcoma	6	6	2	3	.8759	.9050	.9372	.5597
TAIL								
# Evaluated	4	1	3	6				
Adj. # at Risk	3.5	1.0	1.5	3.8				
FIBROSARCOMA	0	0	0	1	.4286	.5000	.	.
Adj. # at Risk	3.5	1.0	1.5	3.8				
HAEMANGIOMA	1	0	0	0	1	1	1	.
UTERINE CERVIX								
# Evaluated	66	64	65	66				
Adj. # at Risk	48.6	44.3	39.8	48.1				
ENDOMETRIAL POLYP	0	0	1	2	.0652	.2474	.4483	.
Adj. # at Risk	48.6	45.2	39.0	47.0				
ENDOMETRIAL STROMAL SARCOMA	0	1	0	0	.7288	.	.	.4839
Adj. # at Risk	48.6	45.2	39.8	48.1				
Endo.Polyp/Stromal Cell Sarc.	0	1	1	2	.1251	.2474	.4483	.4839
Adj. # at Risk	49.2	44.3	39.0	47.0				
LEIOMYOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	48.6	44.9	39.0	47.0				
SCHWANNOMA	0	1	0	0	.7273	.	.	.4783

Reference ID: 3401357

Table A.2.6. (cont.) Tumor Incidence and Results of Tests in Female Mice

Organ/ tumor	Incidence Veh	Low	Med	High	Significance Levels ptrend	phigh vsVeh	pmed vsVeh	plow vsVeh
UTERUS								
# Evaluated	66	66	66	66				
Adj. # at Risk	48.6	45.7	39.3	47.0				
ENDOMETRIAL ADENOCARCINOMA	1	0	1	0	.7284	1	.6985	1
Adj. # at Risk	48.8	46.1	39.3	47.0				
ENDOMETRIAL POLYP	4	3	1	2	.7787	.8882	.9537	.7645
Adj. # at Risk	48.6	45.9	39.7	47.0				
ENDOMETRIAL STROMAL CELL SARCO	0	1	1	0	.6096	.	.4483	.4839
Adj. # at Risk	48.8	46.2	39.7	47.0				
Endo.Polyp/Stromal Cell Sarco	4	4	2	2	.8196	.8882	.8435	.6186
Adj. # at Risk	49.1	45.7	39.4	47.1				
HAEMANGIOMA	2	0	1	1	.5585	.8711	.8321	1
Adj. # at Risk	49.1	45.7	39.3	47.4				
HAEMANGIOSARCOMA	1	0	0	2	.1674	.4842	1	1
Adj. # at Risk	48.7	45.7	39.3	47.0				
LEIOMYOMA	1	1	1	2	.2573	.4839	.6985	.7363
Adj. # at Risk	48.7	45.7	39.3	47.0				
MALIGNANT SCHWANNOMA	1	0	0	0	1	1	1	1
Uterus(w/cervix)								
# Evaluated	66	64	65	66				
Adj. # at Risk	48.6	44.3	39.0	47.0				
ENDOMETRIAL ADENOCARCINOMA	1	0	1	0	.7282	1	.6914	1
Adj. # at Risk	48.8	44.7	39.8	48.1				
ENDOMETRIAL POLYP	4	3	2	4	.4404	.6428	.8435	.7451
Adj. # at Risk	48.6	44.5	39.4	47.0				
ENDOMETRIAL STROMAL CELL SARCO	0	1	1	0	.6111	.	.4483	.4783
Adj. # at Risk	48.6	45.2	39.0	47.0				
ENDOMETRIAL STROMAL SARCOMA	0	1	0	0	.7288	.	.	.4839
Adj. # at Risk	48.6	45.2	39.8	48.1				
Endo.Polyp/Stromal Cell Sarc.	0	1	1	2	.1251	.2474	.4483	.4839
Adj. # at Risk	48.8	44.8	39.4	47.0				
Endo.Polyp/Stromal Cell Sarco	4	4	2	2	.8261	.8882	.8435	.5928
Adj. # at Risk	49.1	44.3	39.0	47.1				
HAEMANGIOMA	2	0	1	1	.5612	.8711	.8262	1
Adj. # at Risk	49.1	44.3	39.0	47.4				
HAEMANGIOSARCOMA	1	0	0	2	.1708	.4842	1	1
Adj. # at Risk	48.7	44.4	39.0	47.0				
LEIOMYOMA	1	1	1	2	.2602	.4839	.6914	.7305
Adj. # at Risk	49.2	44.3	39.0	47.0				
LEIOMYOSARCOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	48.7	44.3	39.0	47.0				
MALIGNANT SCHWANNOMA	1	0	0	0	1	1	1	1
Adj. # at Risk	48.6	44.9	39.0	47.0				
SCHWANNOMA	0	1	0	0	.7273	.	.	.4783

Reference ID: 3401357

Appendix 3. References

Bailer, A. and Portier, C. (1988), "Effects of Treatment-Induced Mortality on Tests for Carcinogenicity in Small Samples", *Biometrics*, **44**, 4, 417-431.

Bieler, G.S., and Williams, R.L. (1993), "Ratio Estimates, the Delta Method, and Quantal Response Tests for Increased Carcinogenicity", *Biometrics*, **49**, 4, 793-801.

Chu, K.C., Ceuto, C., and Ward, J.M. (1981), "Factors in the Evaluation of 200 National Cancer Institute Carcinogen Bioassays", *Journal of Toxicology and Environmental Health,* **8**, 251-280.

Greaves, P. (2007), "Neoplasia of Adrenal Medulla," In: *Histopathology of Preclinical Toxicity Studies.* 3rd edition, pp 819. Oxford, UK: Academic Press, Elsevier Inc.

Haseman, J. K. (1983), "A Reexamination of False-positive Rates for Carcinogenicity Studies", *Fundamental and Applied Toxicology,* **3**, 334-339.

Jara, A. (2007), "Applied Bayesian Non- and Semi-parametric Inference using DPpackage", *Rnews***,** **7**, 3, 17-26.

Lin, K. K. and Ali, M.W. (2006), "Statistical Review and Evaluation of Animal Tumorigenicity Studies", *Statistics in the Pharmaceutical Industry, Third Edition,* edited by C.R. Buncher and J.Y. Tsay, Marcel Dekker, Inc. New York.

Lin, K. K. and Rahman, M.A. (1998), "Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs", *Journal of Biopharmaceutical Statistics*. **8**, 1, 1-15.

McConnell, E.E., Solleveld, H.A., Swenberg, J.A., and Boorman, G.A. (1986), "Guidelines for Combining Neoplasms for Evaluation of Rodent Carcinogenesis Studies", *Journal of the National Cancer Institute*. **76**, 283-289.

Parola, A, and Jacobs, A. (2010). "Combining Tumors for Statistical Analysis", online FDA handout.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). "Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments", *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

Reference ID: 3401357

R Development Core Team (2009). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rahman, M.A. and Lin, K.K. (2008), "A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman", *Journal of Biopharmaceutical Statistics*. **18**, 949-958.

STP Peto Working Group (2002), "Statistical Methods for Carcinogenicity Studies", *Toxicologic Pathology*. **30** (3), 403-414.

U.S. Department of Health and Human Services (2013), Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (DRAFT GUIDANCE), Center for Drug Evaluation and Research, Food and Drug Administration

Reference ID: 3401357

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

/s/

STEVEN F THOMSON
11/04/2013
Statistical Carcinogenicity Review

KARL K LIN
11/05/2013
Concur with review

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 205-677 **Applicant: Vanda** **Stamp Date: 5/31/2013**

Drug Name: Tasimelteon **NDA/BLA Type: priority**

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	x			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	x			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	x			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? __Yes_____

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.				Review issue
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.				Review issue
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			X	
Appropriate references for novel statistical methodology (if present) are included.			x	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.			x	
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.				Review issue

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Jingyu (Julia) Luan, Ph.D.

Reviewing Statistician Date

Kun Jin, Ph.D.

Supervisor/Team Leader Date

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

/s/

JINGYU J LUAN
06/18/2013

KUN JIN
06/18/2013