

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
022416Orig1s000

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 #: 22,416

Drug Name: Eslicarbazepine acetate

Indication(s): Adjunctive therapy in the treatment of partial-onset seizures in adults with epilepsy

Applicant: Sunovion Pharmaceutical Inc.

Date(s): Submission date: 02/11/2013
PDUFA Date: 11/8/13

Review Priority: Priority

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Xiang Ling, Ph.D.

Concurring Reviewers: Kun Jin, Ph.D., Team Leader
James Hung, Ph.D., Director

Medical Division: Division of Neuropharmacological Drug Products, HFD-120

Clinical Team: Teresa Podruchny, M.D.
Mary Doi, M.D.
Norman Hershkowitz, M.D., Ph.D., Team Leader
Eric Bastings, M.D., Acting Division Director

Project Manager: Su-Lin Sun

Table of Contents

1 EXECUTIVE SUMMARY	4
2 INTRODUCTION	5
2.1 OVERVIEW.....	5
2.2 DATA SOURCES	5
3 STATISTICAL EVALUATION	6
3.1 DATA AND ANALYSIS QUALITY.....	6
3.2 EVALUATION OF EFFICACY.....	6
3.2.1 <i>Study Design and Endpoints</i>	6
3.2.2 <i>Statistical Methodologies</i>	7
3.2.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	10
3.2.4 <i>Results and Conclusions</i>	11
3.2.5 <i>Updated Results for Study 301 and Study 302</i>	18
3.3 EVALUATION OF SAFETY	18
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	19
4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION	19
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS.....	20
5 SUMMARY AND CONCLUSIONS	21
5.1 STATISTICAL ISSUES	21
5.2 COLLECTIVE EVIDENCE	21
5.3 CONCLUSIONS AND RECOMMENDATIONS	21

LIST OF TABLES

Table 1. Analysis Populations	10
Table 2. Subject Disposition	10
Table 3. Demographics and Baseline Characteristics	11
Table 4. Primary Analysis of Standardized Seizure Frequency during the Maintenance Period (ITT and DE Population)	12
Table 5. Secondary/Sensitivity Analysis Results for Standardized Seizure Frequency During the Maintenance Period (ITT Population)	13
Table 6. Additional Analyses for Standardized Seizure Frequency (ITT Population)	13
Table 7. Treatment Effect on Standardized Seizure Frequency (Log Difference in LS Mean) by Diary Type and Region	14
Table 8. Responder Analysis	17
Table 9. Relative Change from Baseline in Standardized Seizure Frequency during the Maintenance Period	17
Table 10. Updated Results for Study 301 and Study 302	18
Table 11. Standardized Seizure Frequency by Demographic Subgroups	19
Table 12. Responder Rate and Percent Reduction of Seizure Frequency by Region	20
Table 13. Standardized Seizure Frequency by Baseline Carbamazepine Use	20

1 EXECUTIVE SUMMARY

The data overall provided evidence to support for the efficacy of Eslicarbazepine acetate (ESL) as adjunctive treatment in patients with partial-onset seizure. The ESL 1200 mg dose group was statistically significantly different from placebo with respect to the primary efficacy endpoint (standardized seizure frequency). The ESL 800 mg dose group was not statistically significantly different from placebo, but the results suggested a trend towards an improvement in standardized seizure frequency with this dose. The percent reductions over placebo were 16.3% and 22.9% for ESL 800mg and 1200mg groups respectively.

Study 304 was adequately powered for the analysis of the Daily Entry (DE) diary ITT population. The treatment effects in the DE population were slightly smaller than those of the overall ITT population, and did not achieve statistical significance after correction for multiplicity. There were some issues with the Event Entry (EE) diaries. However, there was not enough evidence to exclude the use of EE diary data. A worst-case type of analysis that excluded subjects using EE diaries still supported the efficacy of ESL 1200 mg dose group.

The discontinuation rate was higher in ESL 1200 mg group with AE being the most common reason for discontinuation. However, as shown in a series of sensitivity analyses, the higher dropout rate did not appear to have a drastic effect on the efficacy results or conclusions.

The effect of ESL was generally consistent across a variety of subgroups defined by demographic and baseline disease characteristics, although there appeared to be some heterogeneity in treatment effect for subgroups by baseline carbamazepine use.

The results from the analyses of the majority of the secondary efficacy endpoints, for example the proportion of responders during the maintenance period, were consistent with the conclusion based on the analysis of the primary efficacy endpoint. The updated results of previous Phase III studies 301 and 302 suggested marginal efficacy of ESL.

2 INTRODUCTION

2.1 Overview

In April 2009, the European Medicines Agency Committee for Medicinal Products for Human Use approved Eslicarbazepine acetate (Zebinix™ and Exalief™) as adjunctive therapy in adults with partial-onset seizures with or without secondary generalization for use in the European Union. The usual maintenance dose is 800 mg once daily and the maximum recommended daily dose is 1200 mg.

Bial was responsible for the initial research and development of Eslicarbazepine acetate (ESL) and Sunovion acquired the rights from Bial in late 2007 to obtain marketing approval and then commercialize ESL in the US and Canadian markets. Sunovion submitted the original New Drug Application (NDA) on 31 March 2009 to the FDA, including two phase III studies (Studies 301 and 302, conducted outside of North America). In April 2010, the FDA issued a Complete Response (CR) Letter, which identified concerns regarding audit findings and the design of the diary cards, as well as discrepancies in clinical data. On 30 July 2010 and 07 June 2011, Sunovion and Bial met with the FDA to discuss the plan for resubmission of the NDA. A new Phase III study 304 was included in this resubmission to address the issues raised in the CR letter. The efficacy endpoints in Study 301 and Study 302 were reanalyzed in the ISE of the resubmission.

All the 3 studies were randomized, placebo-controlled, parallel group, multicenter studies with an 8-week baseline period, a 2-week double-blind titration phase, and a 12-week double-blind maintenance phase. Study 304 used both Daily Entry (DE) diaries, in which subjects entered seizure data every day, irrespective of seizure occurrence (after amendment 3), and Event Entry (EE) diaries, in which subjects entered seizure data when seizures occurred, with no data entered on days without events. Studies 301 and 302 used EE diaries.

2.2 Data Sources

Materials reviewed for this application include the clinical study reports, raw and derived datasets, SAS codes used to generate the derived datasets and tables, protocols, statistical analysis plans, and documents of regulatory communications, which are located in the following directory: <\\Cdsesub1\evsprod\NDA022416\0053\m5>.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Key efficacy endpoints were reproduced by this reviewer from raw data. Documentation of statistical analysis methods was included with sufficient details for this reviewer to reproduce the applicant's key efficacy results. For the previous studies (301 & 302), the reviewer found that extensive hard-codes were used in the creation of the analysis datasets to correct data errors after the database was locked and unblinded, indicating questionable data quality. However, the review did not find it to be the case for Study 304. The quality efficacy data and analyses in Study 304 was acceptable overall.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The original Sponsor of Study 304 was Bial, which planned to conduct the study in Europe and South America. The first subject was enrolled on 02 December 2008. Subsequently, there were a total of 5 protocol amendments, with the final version dated 28 July 2011. On 01 December 2009, the study was expanded to include sponsorship in North America (USA and Canada) by Sunovion (Protocol Amendment No. 2). On 16 September 2010, format of seizure diary from EE to DE diary was implemented for all new enrolled subjects. By then, 168 subjects had been enrolled and they continued to use the EE diary throughout the study. Sample size was increased from 360 to 615 so that the number of subjects using the DE diary design was adequate to support approximately 90% power (Amendment No. 3).

The Statistical Analysis Plan (SAP) was submitted to IND 67,466 serial number 242 on 11 August 2011 for the Agency to review, and was finalized on 06 December 2011. The SAP was amended after final approval as a result of compliance issues identified at Site 952, and was revised and signed off on 05 March 2012. The clinical database was locked and released for unblinding on 12 March 2012. The database was opened and relocked on 25 May 2012 to update the drug accountability data for Subject 00901. Statistical programming and analyses were performed by (b) (4)

Study Design

Study 304 was designed to include 3 parts with Part I supporting efficacy and consisting of an eight-week baseline period, followed by a double-blind two-week titration period and a 12-week maintenance period. Eligible patients would have had at least four partial-onset seizures in the four weeks prior to screening, and a four-week seizure frequency of at least four partial-onset seizures during the eight-week baseline period. At the end of this period, patients who met the selection criteria were randomly assigned in a 1:1:1 ratio to one of the following treatment groups: ESL 800 mg QD, ESL 1200 mg QD or placebo.

The randomization was performed by IVRS following a permuted-block design with a block size of 8, and was stratified by region: North America (USA and Canada) versus Rest of World (ROW). Subjects were to record in their seizure diary all seizures by date, time of occurrence, and seizure type during the study. Overall, of the estimated 615 patients planned to be enrolled, approximately 435 patients would be enrolled using the DE diary. Each site should recruit no more than 18 patients excluding those patients using an EE version of the seizure diary without the Sponsor's approval. The sample size was determined to achieve approximately 90% power for the analysis of DE diary population.

During the entire 12-week maintenance period, dose adjustments in any treatment group were not allowed. Patients who could not tolerate the investigational product at any dose would be withdrawn. Patients who completed Part I would either enter an open-label extension or be tapered off study drug.

Efficacy Endpoints

The primary endpoint was the standardized seizure frequency (per 28 days) over the 12-week maintenance period.

The secondary efficacy endpoints were:

- Proportion of subjects with a 50% or greater reduction in standardized seizure frequency from the baseline period to the 12-week maintenance period (responders).
- Percentage change in standardized seizure frequency during the 2-week titration period, the 12-week maintenance period, and both periods combined.
- Standardized seizure frequency for the titration period and every 28 days during the maintenance periods.
- Percent change from baseline over the maintenance period (< 50%, from 50% through 75% and > 75%, 100%, 25% or greater exacerbation) in standardized seizure frequency.
- Number of seizures occurring during each week.
- Standardized seizure frequency and relative change from baseline (percent) in seizure frequency over the 12-week maintenance period by seizure type.
- Proportion of subjects remaining on treatment for the duration of the study.
- Clinical Global Impressions (CGI) score.
- Seizure Severity Questionnaire (SSQ) overall severity score.
- Quality of Life in Epilepsy Inventory-31 (QOLIE-31) overall score.
- Symptoms of depression assessed by the Montgomery-Asberg Depression Rating Scale (MADRS) total severity score.

3.2.2 Statistical Methodologies

Efficacy Analysis Population

The ITT population is the primary population for the analysis of efficacy, consisting of all randomized subjects who received at least one dose of study drug after randomization and had at least one post-baseline seizure frequency assessment. This population includes subjects who completed the EE diary and subjects who completed the DE diary.

DE Diary ITT population – all subjects in the ITT population who completed the DE diary.

EE Diary ITT population – all subjects in the ITT population who completed the EE diary.

Per-protocol (PP) population – all patients in the ITT population who did not have any major important protocol deviations.

Multiplicity Adjustment

To control the family-wise Type I error for the analysis of the primary efficacy variable in the ITT and DE Diary ITT populations, a 2-stage gate-keeping procedure was used (Dmitrienko et al *Biometrical Journal* 2008). In Stage 1, each of the two pair-wise comparisons of ESL 800 mg and 1200 mg to placebo in the ITT population was conducted using an alpha level of 0.025. If both comparisons were statistically significant, the Dunnett's method would be conducted at an alpha level of 0.05 for the DE population in Stage 2. If only one of two pair-wise comparisons was statistically significant in Stage 1, an alpha level of 0.025 would be carried over to the Stage 2 testing. If neither of the two pair-wise comparisons was statistically significant in Stage 1, the procedure would stop and no further analyses would be performed for the DE Diary ITT population.

Analysis of the Primary Endpoint

For the purpose of the primary efficacy analysis, standardized seizure frequency was natural logarithmically transformed (ln). To avoid the ln(0) problem, a constant (0.333) was added. The ln standardized seizure frequency during the maintenance period was analyzed using an ANCOVA model. The model included treatment as a fixed effect, ln standardized seizure frequency at baseline, and diary version as covariates. The treatment estimates (LSMeans and 95% CIs) were back-transformed using the exponential function and subtracting 0.333, and SEs were calculated using the Delta Method.

The interaction between treatment and diary version was tested in a separate ANCOVA model for the ITT population. The statistical significance of the interaction was assessed at an alpha level of 0.10. The assumptions of the ANCOVA models were checked using residual-by-predicted value plots and normality probability plots. The sponsor pre-specified extensive secondary/sensitivity analyses for the primary endpoint. The reviewer also conducted a sensitivity analysis using non-parametric ANCOVA.

Analyses to Assess the Impact of Dropouts

If subjects discontinued during the 2-week titration period, before the start of the 12-week maintenance period, they were treated as missing in the primary efficacy analysis. The following analyses were conducted to assess the impact of early dropouts.

1. Missing standardized seizure frequency during the maintenance period was imputed using subjects' standardized seizure frequency during the titration period.

2. Analysis based on using the seizure frequency during the baseline period as the post-discontinuation frequency (i.e, seizure frequency between the date of discontinuation/last non-missing diary data and the scheduled end of the maintenance period). The standard seizure frequency for the maintenance period was then calculated using average of the observed pre-discontinuation frequencies and the imputed post-discontinuation frequencies weighted according to the number of days.
3. Analysis based on using the seizure frequency during the last two weeks prior to discontinuation as the post-discontinuation frequency. The last two weeks prior to discontinuation can be in any of the study periods, for example, if a patient only has one week of seizure frequency data during the titration period, then the seizure frequency during the last week of the baseline period will be used in the imputation.

The reviewer conducted the following additional analyses:

- a. Using seizure frequency for the combined titration and maintenance period.
- b. Using seizure frequency for last 14 days for subjects with less than 14 days of seizure data during the maintenance period (no imputation for subjects who dropped out but had at least 14 days of seizure data during the maintenance period).

Subgroup Analyses of the Primary Endpoint

The primary efficacy variable was also analyzed using ANCOVAs that model ln standardized seizure frequency during the maintenance period as a function of ln standardized seizure frequency at baseline, diary version, treatment, and each of the following covariates and their interactions with treatment:

- Region (North America versus ROW);
- Age (<40 years, 40-65 years, >65 years);
- Race (Caucasian, non-Caucasian);
- Sex (Male, Female);
- Most Common AED Use during Baseline Period;
- Carbamazepine (CBZ) Dose Reduction during Maintenance Period (Yes, No);
- Phenytoin Dose Reduction during Maintenance Period (Yes, No);
- Use of Rescue Medication during Maintenance Period (Yes, No).

Analyses of the Secondary Endpoints

For the proportion of patients who were responders, seizure-free, exacerbated, and distribution of seizure reduction, CMH test for ordinal data were used, stratified by region and diary version (ITT and PP populations) or by region only (EE Diary ITT and DE Diary ITT populations).

The relative (percentage) change in standardized seizure frequency during the maintenance period was analyzed as described above for the primary efficacy endpoint. An additional non-parametric ANCOVA based on ranked data was also conducted for the ITT population.

Subjects who did not have maintenance data were not included in those analyses.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

Part I of Study 304 was conducted in 19 countries at 173 sites which screened 936 subjects. Of these, 653 subjects were randomized at 160 sites. There were 640 subjects included in the ITT population, consisting of 185 subjects using EE diaries and 455 subjects using DE diaries (Table 1).

Table 1. Analysis Populations

Data Analysis Sets	Placebo N=226 n (%)	ESL 800 mg N=216 n (%)	ESL 1200 mg N=211 n (%)	Total Randomized N=653 n (%)
ITT Population	220 (97.3)	215 (99.5)	205 (97.2)	640 (98.0)
EE ITT Population	62 (27.4)	67 (31.0)	56 (26.5)	185 (28.3)
DE ITT Population	158 (69.9)	148 (68.5)	149 (70.6)	455 (69.7)
PP Population	188 (83.2)	184 (85.2)	175 (82.9)	547 (83.8)

Source: Table 10 of the CSR for Study 304.

The Part I completion rate was 83.6%, 80.1%, and 67.3% for the placebo, ESL 800 mg and ESL1200 mg treatment groups, respectively. A greater number of discontinuations was observed in the ESL 1200 mg group. The most common reason for premature discontinuation was AEs for all treatment groups, with 4.0%, 9.7%, and 21.3% in the placebo, ESL 800 mg and ESL 1200 mg groups, respectively (Table 2).

Table 2. Subject Disposition

Disposition	Placebo n (%)	ESL 800 mg n (%)	ESL 1200 mg n (%)	Total n (%)
Randomized	226 (100.0%)	216 (100.0%)	211 (100.0%)	653 (100.0%)
Randomized and Received at Least 1 Dose of Study Drug (Safety Population)	224 (99.1%)	216 (100.0%)	210 (99.5%)	650 (99.5%)
Entered the Maintenance Period	212 (93.8%)	201 (93.1%)	184 (87.2%)	597 (91.4%)
Completed the Double-Blind Period	189 (83.6%)	173 (80.1%)	142 (67.3%)	504 (77.2%)
Prematurely Discontinued from the Double-Blind Period	37 (16.4%)	43 (19.9%)	69 (32.7%)	149 (22.8%)
Entered the 1-Year Open-Label Period	187 (82.7%)	171 (79.2%)	140 (66.4%)	498 (76.3%)
Primary Reason for Discontinuation from the Double-Blind Period ^a				
Administrative Reasons	1 (0.4%)	2 (0.9%)	1 (0.5%)	4 (0.6%)
Adverse Event	9 (4.0%)	21 (9.7%)	45 (21.3%)	75 (11.5%)
Lack of Efficacy	0	0	1 (0.5%)	1 (0.2%)
Non-Compliance with Study Drug	5 (2.2%)	1 (0.5%)	3 (1.4%)	9 (1.4%)
Physician Decision	1 (0.4%)	0	3 (1.4%)	4 (0.6%)
Pregnancy	2 (0.9%)	1 (0.5%)	0	3 (0.5%)
Protocol Violation	4 (1.8%)	3 (1.4%)	3 (1.4%)	10 (1.5%)
Withdrawal by Subject	7 (3.1%)	7 (3.2%)	12 (5.7%)	26 (4.0%)
Other	8 (3.5%)	8 (3.7%)	1 (0.5%)	17 (2.6%)

Source: Table 9 of the CSR for Study 304.

The overall demographic data and baseline disease characteristics were summarized in Table 3. The treatment groups were balanced for age, sex, and race. Overall, the number of AEDs being taken at baseline was similar across the treatment groups. The majority of subjects were taking 2 AEDs at baseline. The treatment groups were comparable with respect to standardized seizure frequency during the baseline period (Table 3).

Best
Available
Copy

Table 3. Demographics and Baseline Characteristics

Characteristic	Placebo (N = 224)	ESL 800 mg (N = 216)	ESL 1200 mg (N = 210)
Age (years)			
n	223	216	210
Mean	39.0	38.8	38.0
SD	12.70	12.11	11.98
Median	39.0	38.5	38.0
< 40 years	115 (51.6%)	113 (52.3%)	115 (54.8%)
40 – 65 years	105 (47.1%)	100 (46.3%)	93 (44.3%)
> 65 years	3 (1.3%)	3 (1.4%)	2 (1.0%)
Missing	1	0	0
Sex, n %			
Male	112 (50.0%)	109 (50.5%)	105 (50.0%)
Female	112 (50.0%)	107 (49.5)	105 (50.0%)
Race, n %			
White	142 (63.4%)	137 (63.4%)	134 (63.8%)
Black or African American	8 (3.6%)	8 (3.7%)	8 (3.8%)
Asian	46 (20.5%)	41 (19.0%)	39 (18.6%)
Other	28 (12.5%)	30 (13.9%)	29 (13.8%)
Number of AEDs at Baseline			
1 AED	64 (28.6%)	60 (27.8%)	59 (28.1%)
2 AEDs	158 (70.5%)	153 (70.8%)	151 (71.9%)
3 or more AEDs	1 (0.4%)	0	0
AEDs during the Baseline Period Used by > 15% of Subjects			
Carbamazepine	77 (34.4%)	84 (38.9%)	89 (42.4%)
Levetiracetam	66 (29.5%)	58 (26.9%)	43 (20.5%)
Lamotrigine	57 (25.4%)	51 (23.6%)	57 (27.1%)
Valproic acid	42 (18.8%)	46 (21.3%)	41 (19.5%)
Baseline Standardized Seizure Frequency (ITT population)			
n	220	215	204
Mean	16.3	18.2	17.2
SD	19.29	34.49	21.08
Median	9.0	8.6	8.9

Source: Table 12, 13 & 15 of the CSR for Study 304.

3.2.4 Results and Conclusions

3.2.4.1 Analyses of the Primary Endpoint

The study utilized a two stage gate-keeping multiple test procedure (MTP) for type I error control. In the first stage, a Bonferroni adjustment was applied and each of the comparisons of the ESL 800 mg and ESL 1200 mg treatment groups to placebo in the ITT population was conducted at an alpha level of 0.025. There was a statistically significant difference between the ESL 1200 mg treatment group and placebo treatment group (adjusted p-value=0.004); the difference between the placebo and the ESL 800 mg treatment group did not reach statistical

significance (adjusted p-value =0.058). The percent reductions over placebo calculated by $100*[1-\exp(\text{LSMean difference of the log standardize seizure frequency})]$ were 16.3% and 22.9% for ESL 800mg and 1200mg groups respectively. The dose response appeared to be monotone.

In the second stage, Dunnett MTP at alpha level of 0.025 was used for the two comparisons in the DE ITT population. The adjusted p-values for the second step analysis of the DE diary population indicated that the difference from placebo was not statistically significant for either ESL treatment groups at the 0.025 level (Table 4).

Table 4. Primary Analysis of Standardized Seizure Frequency during the Maintenance Period (ITT and DE Population)

	Placebo	ESL 800 mg	ESL 1200 mg
ITT population			
N ^a	212	200	184
LS mean (SE)	7.88 (0.49)	6.54 (0.41)	6.00 (0.40)
95% CI	[6.98, 8.90]	[5.77, 7.40]	[5.26, 6.84]
Log difference in LS mean		-0.18	-0.26
Unadjusted p-value		0.029	0.002
Adjusted p-value ^b		0.058	0.004
DE ITT population			
N ^a	154	137	136
LS mean (SE)	7.54 (0.54)	6.32 (0.48)	5.96 (0.46)
95% CI	[6.55, 8.68]	[5.44, 7.35]	[5.12, 6.94]
Log difference in LS mean		-0.17	-0.22
Unadjusted p-value		0.094	0.026
Adjusted p-value ^c		0.167	0.049

a Subjects who discontinued from the study during the titration period were not included.

b Bonferroni's procedure was used to calculate the p- values.

c Dunnett's procedure was used to calculate the p-values (assessed at 0.025 level).

Source: Table 19 & 20 of the CSR for Study 304, confirmed by the reviewer.

The reviewer conducted a sensitivity analysis on the ITT population using non-parametric ANCOVA. The result was consistent with the primary analysis. The unadjusted p-values were 0.038 for the ESL800 mg group and 0.006 for the ESL1200 mg group.

3.2.4.2 Assessment of the Impact of Dropouts on the Primary Analysis Result

In the primary analyses, subjects who discontinued from the study during the titration period were not included. The pre-specified secondary/sensitivity analyses (Table 5) and the reviewer's additional analyses (Table 6) assessed the impact of early dropouts. Note that there were more subjects in high dose group withdrew in early treatment phase, and those subjects tended to under-report seizure events hence had lower seizure frequency during the titration phase. Therefore, the results based one titration period observation carried forward and combined titration and maintenance period may overestimate the treatment effect and result in smaller p-values.

For example, of the 27 DE subjects who dropped out during the titration phase, 8 subjects (30%) had diary compliance (defined as the number of days of diary data completed/number of days of diary data expected to be completed x 100) less than 30%. For subjects using EE diary, seizure data was considered missing only when diary card had not been returned. However, similar to the DE diaries, it was possible that some seizure data were missing even when the diary card had been returned. Those missing EE diary data cannot be known for certain and was not accounted for in the analysis.

The imputations using the last 2-week seizure frequency prior to dropout seemed more sensible and the results were consistent with that of the primary analysis.

Table 5. Secondary/Sensitivity Analysis Results for Standardized Seizure Frequency During the Maintenance Period (ITT Population)

	Placebo N=220	ESL 800 mg N=215	ESL 1200 mg N=205
Titration Period Observations Carried Forward			
LS mean (SE)	8.12 (0.54)	6.18 (0.42)	5.82 (0.41)
Log difference in LS mean		-0.26	-0.32
Unadjusted p-value		0.003	<0.001
Imputation with Baseline Frequency			
LS mean (SE)	8.16 (0.44)	7.03 (0.38)	6.80 (0.39)
Log difference in LS mean		-0.14	-0.17
Unadjusted p-value		0.045	0.016
Imputation with Last 2-week Seizure Frequency			
LS mean (SE)	8.06 (0.50)	6.70 (0.42)	5.80 (0.38)
Log difference in LS mean		-0.18	-0.31
Unadjusted p-value		0.030	<0.001

Source: Table 21 & 14.2.15.3 & 14.2.15.4 of the CSR for Study 304, confirmed by the reviewer.

Table 6. Additional Analyses for Standardized Seizure Frequency (ITT Population)

	Placebo N=220	ESL 800 mg N=215	ESL 1200 mg N=205
Combined Titration and Maintenance Period			
LS mean (SE)	8.68 (0.52)	6.60 (0.40)	6.31 (0.39)
Log difference in LS mean		-0.26	-0.31
Unadjusted p-value		0.001	<0.001
Imputation with Last 2-week Seizure Frequency for Subjects with <14 days of Maintenance Seizure Data			
LS mean (SE)	8.07 (0.50)	6.67 (0.42)	5.87 (0.38)
Log difference in LS mean		-0.18	-0.30
Unadjusted p-value		0.025	<0.001

Source: The FDA reviewer.

3.2.4.3 Analyses Concerning the Diary Format and Region

The study was expanded to North America after Protocol Amendment No. 2, and the format of seizure diary was changed from EE to DE diary was after Protocol Amendment No. 3. There were no statistically significant interactions between treatment and diary version, or between treatment and region at 0.10 level. However, EE diary ITT population had numerically larger treatment effect compared to DE diary ITT population, and the treatment effect in the Rest of the World (ROW) were larger compared to North America. The largest treatment effect was seen in subjects using EE diary in the ROW (Table 7). This was the population in the originally protocol of Study 304, as well as in the previous Phase III Studies 301 & 302.

Table 7. Treatment Effect on Standardized Seizure Frequency (Log Difference in LS Mean) by Diary Type and Region

	North America			Rest of World			Total		
		ESL 800 mg	ESL 1200 mg		ESL 800 mg	ESL 1200 mg		ESL 800 mg	ESL 1200 mg
EE	N=44	-0.10	-0.16	N=125	-0.34	-0.46	N=169	-0.21	-0.36
DE	N=168	-0.05	-0.17	N=260	-0.24	-0.26	N=428	-0.17	-0.22
Total	N=212	-0.03	-0.18	N=385	-0.27	-0.32	N=597	-0.18	-0.26

N is the number of subject with maintenance seizure data for each subgroup.

Source: The FDA reviewer.

In North America, there were few subjects using EE diary, and the placebo response seemed to be larger compared to ROW (see section 4.2). In the ROW, there was significant effect of Diary Type (p-value=0.01), suggesting that Diary Type had an effect on the seizure frequency regardless of the treatment. The p-value of the Diary Type was 0.06 for the overall ITT population. In the review of Studies 301 & 302 for the original NDA, this reviewer pointed out the deficiency of the EE diaries: as subjects were instructed to update seizure diary only when they experienced a seizure, true zero seizure could not be differentiated from missing seizure data. Based on the Agency's comments, DE diary was utilized for new patients in Study 304 at the time of Amendment No. 3. In addition, for subjects using the EE diary, a seizure diary tracking log was completed by the sites (retrospectively and prospectively). The end of a seizure evaluation period was defined as the last date that the patient returned their diary. If there were missing diaries for patients during an evaluation period, the number of days in which diaries were missing was not included in the calculation of the average daily frequency for that period. Zero seizures were only assumed in cases where the diary card had been returned. Although those measures attempted to overcome the deficiencies of the EE diaries, there were still some issues concerning the EE diaries as described below.

1. For subjects using EE diaries, when no seizures were reported, their seizure data were considered as zeros if the dates were covered by the presence of a corresponding diary card on the Diary Tracking Log CRF. However, even when the diary card had been returned, it was possible that some seizure data were missing. This type of missing data

cannot be identified due to the limitation of EE diary, and cannot be accounted for in the analysis. On the contrary, missing seizure data were captured in DE diaries. Overall, a high level of compliance was observed for DE diary. The majority of subjects (77%, 80%, and 80% for the placebo, ESL 800 mg, and ESL 1200 mg groups, respectively) missed no days or only 1 day. The days with missing seizure data were excluded in the calculation of seizure frequency.

2. For subjects using EE diaries, seizure data were considered missing if the dates were not covered on the Diary Tracking Log CRF by the dates that the EE diary cards were dispensed and returned. For example, subject 20101 had 32 seizures recorded on 28 days between 19 February 2009 and 15 April 2009 (8-week baseline period). However, the diary tracking log did not contain records of diaries dispensed or returned during this period. Therefore, days without seizures during this period (February 22, February 24, etc., see the CRF below) were set to be missing. In the reviewer's opinion, it was likely that the subject did not have any seizure on those days and seizure data was set to missing due to the error in the diary tracking log. However, only 3 subjects had similar situation and the impact on the efficacy result was minimal.

(b) (4)

3. The sites were instructed to transcribe the seizure records on the EE diaries onto CRF pages and errors could occur during this manual process. For example, the seizures reported by subject #00405 between 29 July 2010 and 4 August 2010 were transcribed twice to the CRF, therefore, there were duplicate seizures in the dataset and those seizures were double counted in the calculation of seizure frequency. The Agency

requested the sponsor to audit a total of 40 patients' diary and database, including subjects #00405 and other random selected patients who potentially had duplicate seizure records. The sponsor found out some duplicate seizure records, but failed to identify the problem for subject #00405.

4. For subjects using EE diaries, the end of a seizure evaluation period was defined as the last date that the patients returned their diary. This is a reasonable assumption, but there may be exceptions. For subject #00405, the last seizure reported on CRF was on 12 September 2010, and the diary tracking log showed that the last diary was returned on 12 October 2010. Therefore, it was assumed that no seizure occurred during the 30 days between September 12 and October 12. However, this subject had 169 seizures during the 80 days prior to September 12. It was questionable that whether this subject really did not experience any seizures or he/she just did not record the seizures. The extent of this problem could not be known for certain, as those zero seizures may or may not be true.

Based on the review of the dataset and select CRFs, the problems noted above were not deemed common. There could be other problems that we have not identified yet. It is not sure if collectively they could undermine the credibility of EE diary. However, the evidence to date may not be enough to dismiss EE diary data entirely, although some sort of discounting of the EE diary data may be reasonable. As a worst case analysis in which EE diary data were excluded, the result of the analysis base on only DE diary ITT population showed a statistically significant difference between the ESL 1200 mg treatment group and placebo treatment group (adjusted p-value=0.049, Table 4). A significance level of 0.05 should be used in this case since the EE diary was excluded and the analysis of the DE diary ITT was treated as the primary analysis.

3.2.4.4 Analyses of Covariates

Analyses with the additional covariates for the primary endpoint were performed. No significant effects were observed for age group, region or race. The following covariates were found to be statistically significant in the ANCOVA analyses: sex, baseline carbamazepine use, baseline lamotrigine use, and baseline valproic acid use (p-value \leq 0.05). None of the treatment-by-covariate interactions (other than treatment-by-carbamazepine use interaction) were statistically significant. Results of subgroup analyses were presented in Section 4.

3.2.4.5 Analyses of Secondary Endpoints

Subjects who had at least a 50% reduction from baseline in standardized seizure frequency during the maintenance period were classified as responders. The sponsor's analysis excluded subjects without data during maintenance period. The reviewer did a sensitivity analysis in which subjects without maintenance data were considered non-responders. The results were similar. Based on the reviewer's analysis, the overall percentage of responders was 22.3% in the placebo group, 28.4% in the ESL 800 mg group, and 30.1% in the ESL 1200 mg group. The unadjusted p-values for the difference from placebo were 0.123 for the ESL 800 mg group and <0.001 for the ESL 1200 mg group (Table 8).

Table 8. Responder Analysis

	Placebo (N = 220)	ESL 800 mg (N = 215)	ESL 1200 mg (N = 205)
Subjects without maintenance data were excluded			
Yes	49 (23.1%)	61 (30.5%)	78 (42.6%)
No	163 (76.9%)	139 (69.5%)	105 (57.4%)
Not Evaluable	8	15	22
Exact 95% CI for Percentage of Responders	(17.6%, 29.4%)	(24.2%, 37.4%)	(35.4%, 50.1%)
p-value		0.068	< 0.001
Subjects without maintenance data were considered non-responders			
Yes	49 (22.3%)	61 (28.4%)	78 (30.1%)
No	171 (77.7%)	154 (71.6%)	127(62.0%)
Exact 95% CI for Percentage of Responders	(16.8%, 27.8%)	(22.4%, 34.4%)	(31.4%, 45.1%)
p-value		0.123	< 0.001

Source: FDA reviewer.

Table 9 presented the relative change from baseline during the maintenance period for the ITT population. The median percentage change from baseline was -21.8% in the placebo group, -9.7% in the ESL 800 mg group, and -35.6% in the ESL 1200 mg group. The unadjusted p-values from the parametric analysis were 0.074 for ESL 800 mg versus placebo and 0.021 for ESL 1200 mg versus placebo. For the non-parametric analysis based on ranked data, the sponsor ranked the relative change in standardized seizure frequency and the baseline standardized seizure frequency values within each treatment group, while the reviewer ranked each of the variables across the treatment groups. Both achieved statistical significance for ESL 1200 mg group.

Table 9. Relative Change from Baseline in Standardized Seizure Frequency during the Maintenance Period

Parameter	Placebo (N = 220)	ESL 800 mg (N = 215)	ESL 1200 mg (N = 205)
n	212	200	183
Mean	-8.7	-23.5	-28.6
SD	121.5	53.9	51.9
25 th Percentile	-47.2	-55.4	-66.9
Median	-21.8	-29.7	-35.6
75 th Percentile	3.9	6.3	-1.7
Parametric Analysis			
LS mean (SE)	-7.14 (6.04)	-21.99 (6.11)	-26.73 (6.48)
Log difference in LS mean		-14.84	-19.59
Unadjusted p-value	-	0.074	0.021
Non-Parametric Analysis			
Unadjusted p-value (Sponsor's analysis)		0.249	0.012
Unadjusted p-value (Reviewer's analysis)		0.068	0.004

Source: Table 14.2.5.1.1 of the CSR for Study 304 and FDA reviewer.

3.2.5 Updated Results for Study 301 and Study 302

The efficacy variables in Study 301 and Study 302 were reanalyzed in the ISE of this resubmission. Re-analysis was performed to account for the effect of seizure diaries not previously analyzed, exclusion of two non-compliant sites, and redefinition of study periods. For Study 301, two sites (301-174 and 301-175 with a total of 20 subjects) were excluded from the ISE re-analyses due to GCP violation (original diary cards were not maintained). An extensive audit program conducted after the original NDA identified additional seizure diary pages which were omitted from the study database for seven subjects from Study 301 and one subject from Study 302. In the original CSR analysis, subjects in Study 301 and Study 302 were evaluated according to the length of time they participated in the study; if no seizures were reported for a day, it was assumed that no seizure had occurred while the subject was still in the study. In the re-analysis, the last diary card return date was utilized to cap the end of the study period.

The updated results were in Table 10. Although the updated results still reached statistical significance for both ESL800 mg dose and ESL1200 mg dose, the monotone dose-response was seen only in Study301 but not in Study302, and the significance of ESL 800 mg group in Study 301 and ESL 1200 mg group in Study 302 became marginal. The p-value for the ESL 800 mg group in Study 301 changed from 0.003 in the original NDA to 0.047 in ISE of the resubmission, and the p-value for the ESL 1200 mg group in Study 302 changed from 0.001 to 0.042.

Table 10. Updated Results for Study 301 and Study 302

	Placebo	ESL 400 mg	ESL 800 mg	ESL 1200 mg
Study 301				
N	95	91	88	87
LS mean (SE)	6.6 (0.54)	5.8 (0.48)	5.0 (0.43)	4.3 (0.38)
Adjusted p-value	-	0.4969	0.0468	0.0010
Study 302				
N	99	94	87	81
LS mean (SE)	8.6 (0.62)	8.1 (0.60)	6.2 (0.48)	6.6 (0.53)
Adjusted p-value	-	0.9043	0.0057	0.0424

Source: Table 31 of ISE.

3.3 Evaluation of Safety

Please see the clinical review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The analysis results for the primary endpoint by demographic subgroups were in Table 11. The treatment effect was generally consistent across the subgroups, although it seemed a bit greater in male subjects. LS mean standardized seizure frequencies were lower in each ESL group as compared to the placebo group. A dose-dependent trend was observed for all the subgroups except for the group of age < 40 years. There were few subjects aged > 65 years in this study; therefore, no statistical comparisons were performed in this age category.

Table 11. Standardized Seizure Frequency by Demographic Subgroups

	Placebo	ESL 800 mg	ESL 1200 mg
Sex: Female			
N	106	99	92
LS mean (SE)	8.83 (0.73)	7.86 (0.66)	6.97 (0.62)
Log difference in LS mean		-0.11	-0.23
Unadjusted p-value		0.311	0.045
Sex: Male			
N	106	101	92
LS mean (SE)	6.96 (0.64)	5.38 (0.51)	5.07 (0.52)
Log difference in LS mean		-0.24	-0.30
Unadjusted p-value		0.043	0.015
Race: Caucasian			
N	137	127	114
LS mean (SE)	7.87 (0.56)	6.66 (0.50)	6.27 (0.50)
Log difference in LS mean		-0.16	-0.22
Unadjusted p-value		0.100	0.031
Race: Non-Caucasian			
N	75	73	70
LS mean (SE)	8.03 (0.98)	6.39 (0.75)	5.60 (0.68)
Log difference in LS mean		-0.22	-0.34
Unadjusted p-value		0.145	0.024
Age: < 40 years			
N	109	103	106
LS mean (SE)	8.06 (0.77)	6.50 (0.62)	6.58 (0.63)
Log difference in LS mean		-0.21	-0.19
Unadjusted p-value		0.095	0.112
Age: 40-65 years			
N	99	94	76
LS mean (SE)	7.68 (0.62)	6.64 (0.56)	5.54 (0.53)
Log difference in LS mean		-0.14	-0.31
Unadjusted p-value		0.203	0.008

Source: Tables 14.2.14.2.1, 14.2.14.2.2 & 14.2.14.2.3 of the CSR for Study 304.

The analysis results for the primary endpoint by region subgroups were shown in Table 7. Table 12 presented the responder rate and percent reduction of seizure frequency by region. The percentages of responders and percent reduction of seizure frequency were similar in the North America and the ROW for the ESL 800 mg and ESL 1200 mg groups, but appeared greater in the North America than the ROW for the placebo group.

Table 12. Responder Rate and Percent Reduction of Seizure Frequency by Region

	Placebo	ESL 800 mg	ESL 1200 mg
North America			
N	75	70	65
Responder rate n(%)	21 (28.0%)	21 (30.0%)	26 (40.0%)
Median Percent Reduction (%)	-25.0	-29.7	-35.6
ROW			
N	142	138	131
Responder n(%)	28 (20.4%)	40 (30.8%)	52 (44.1%)
Median Percent Reduction (%)	-15.3	-28.3	-35.1

Source: FDA reviewer.

4.2 Other Special/Subgroup Populations

There appears to be some heterogeneity in treatment effect for subgroups by baseline carbamazepine use (Table 13).

Table 13. Standardized Seizure Frequency by Baseline Carbamazepine Use

	Placebo	ESL 800 mg	ESL 1200 mg
Baseline Carbamazepine Use: No			
N	144	132	107
LS mean (SE)	8.29 (0.67)	6.55 (0.57)	5.14 (0.48)
Log difference in LS mean		-0.22	-0.46
Unadjusted p-value		0.038	<0.001
Baseline Carbamazepine Use: Yes			
N	73	77	77
LS mean (SE)	7.23 (0.67)	6.51 (0.57)	7.35 (0.67)
Log difference in LS mean		-0.10	0.02
Unadjusted p-value		0.407	0.889

Source: Tables 14.2.14.2.4 of the CSR for Study 304.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

The discontinuation rate was higher in ESL 1200 mg group with AE being the most common reason for discontinuation. However, as shown in a series of sensitivity analyses, the higher dropout rate did not appear to have a drastic effect on the efficacy results or conclusions.

Study 304 was adequately powered for the analysis of the DE diary ITT population. The treatment effects in the DE population were slightly smaller than those of the overall ITT population, and did not achieve statistical significance after correction for multiplicity. There were some issues with the EE diary. However, there was not enough evidence to exclude the use of EE diary data. A worst-case type of analysis that excluded subjects using EE diaries still supported the efficacy of ESL 1200 mg dose group.

5.2 Collective Evidence

The ESL 1200 mg dose group was statistically significantly different from placebo with respect to the primary efficacy endpoint (standardized seizure frequency). The ESL 800 mg dose group was not statistically significantly different from placebo, but the results suggest a trend towards an improvement in standardized seizure frequency with this dose. The percent reductions over placebo were 16.3% and 22.9% for ESL 800mg and 1200mg groups respectively.

The effect of ESL was generally consistent across a variety of subgroups defined by demographic and baseline disease characteristics, although there appeared to be some heterogeneity in treatment effect for subgroups by baseline carbamazepine use.

The results from the analyses of the majority of the secondary efficacy endpoints, for example the proportion of responders during the maintenance period, were consistent with the conclusion based on the analysis of the primary efficacy endpoint. The updated results of previous Phase III studies 301 and 302 suggested marginal efficacy of ESL.

5.3 Conclusions and Recommendations

The data overall provided evidence to support for the efficacy of Eslicarbazepine acetate as adjunctive treatment in patients with partial-onset seizure.

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

/s/

XIANG LING
09/09/2013

KUN JIN
09/10/2013
I concur with the review.

HSIEN MING J HUNG
09/10/2013



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

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number: 022-416

Drug Name: Eslicarbazepine Acetate

Indication: Epilepsy (adjunct therapy of partial onset seizures in adults)

Study number: Study 093-153

Applicant: Sunovion Pharmaceuticals Inc.

Date(s): Date of Document: 2/10/2013
PDUFA date: 8/10/2013
Completion date: 5/15/13

Review Priority: P

Biometrics Division: DB VI

Statistical Reviewer: Ling Chen, Ph.D., Mathematical Statistician, Special Projects Team, DBVI/OB/OTS

Concurring Reviewers: Stella Machado, Ph.D., Division Director, and Acting Team Leader, DBVI/OB/OTS

Medical Division: Controlled Substance Staff

The CSS Team: Alicja Lerner, M.D., Ph.D., Medical Officer, OD/CSS
Katherine R. Bonson, Ph.D., Pharmacologist, OD/CSS
Michael Klein, Ph.D., Director, OD/CSS

Project Manager: Sandra Saltz, OD/CSS

Keywords: Crossover design; Drug abuse potential study; Self-reported endpoint; Multiple endpoints

Table of Contents

LIST OF TABLES.....	3
LIST OF FIGURES.....	3
1. EXECUTIVE SUMMARY	4
2. REVIEW REPORT ON STUDY 093-153	5
2.1 OVERVIEW	5
2.1.1 <i>Objectives of the study</i>	5
2.1.2 <i>Study design</i>	5
2.1.3 <i>Abuse potential measure and data collection times</i>	6
2.1.4 <i>Number of subjects</i>	7
2.1.5 <i>Statistical methodologies used in the Sponsor’s analyses</i>	7
2.1.5 <i>Sponsor’s Summary and Conclusions</i>	8
2.2 DATA LOCATION	10
2.3 REVIEWER’S ASSESSMENT.....	10
2.3.1 <i>Missing data issue</i>	10
2.3.2 <i>Primary Analysis</i>	13
2.3.2.1 <i>Descriptive Statistics</i>	13
2.3.2.2 <i>Statistical Testing</i>	17
2.3.3 <i>Secondary Analysis</i>	18
3. CONCLUSION	22
4. REFERENCE	23
5. APPENDICES	24
5.1 APPENDIX I.....	24
5.2 APPENDIX II.....	26

List of Tables

Table 1: Summary statistics for Emax of Drug Liking VAS (N=44)	14
Table 2: Statistical analysis results for Drug Liking VAS	18
Table 3: Statistical analysis results for ARCI MBG and High VAS	18
Table 4: Summary statistics and testing results for Emax of ARCI MBG	19
Table 5: Summary statistics for Emax of ARCI PCAG	20
Table 6: Summary statistics and testing results for Emax of Good Effects VAS	20
Table 7: Summary statistics for Emax of High VAS	21
Table 8: Summary of the results from significance tests for the abuse potential measures considered in this review	22
Table 9: P-values of the significance tests for Emax of ARCI MBG (n=44)	26
Table 10: P-values of the significance tests for Emax of Good Effects (n=44)	26

List of Figures

Figure 1: Individual time course response profiles for Drug Liking VAS (A3)	11
Figure 2: Individual time course response profiles for Drug Liking VAS (A1.5)	11
Figure 3: Individual time course response profiles for Drug Liking VAS (E2400)	12
Figure 4: Individual time course response profiles for Drug Liking VAS (A2 in Qualification Phase)	13
Figure 5: Boxplots for seven treatments and the differences between eslicarbazepine acetate and alprazolam, and between eslicarbazepine acetate and placebo for Drug Liking VAS (N=44)	15
Figure 6: Mean and median dose response curves for Drug Liking VAS (N=44)	15
Figure 7: Dose response curves of eslicarbazepine acetate in difference from alprazolam and placebo for Drug Liking VAS (n=44)	16
Figure 8: Mean time course profiles for Drug Liking VAS (N=44)	16
Figure 9: Heat map for Emax of Drug Liking VAS by treatment	17
Figure 10: Mean time course profiles for ARCI MBG	24
Figure 11: Mean time course profiles for ARCI PCAG	24
Figure 12: Mean time course profiles for Good Effects VAS	25
Figure 13: Mean time course profiles for High VAS	25

1. Executive Summary

Study 093-153 was a randomized, double-blind, placebo- and active-controlled 7-way crossover study to evaluate the abuse potential of single doses (80 mg, 1600 mg, 2000 mg, and 2400 mg) of eslicarbazepine acetate compared to placebo and 2 doses (1.5 mg and 3 mg) of alprazolam in recreational CNS depressant users.

Of the 53 subjects who were randomized to the Treatment Phase, 49 subjects (92.5%) were included in the PK Population and 44 subjects (83.0%) completed the study. The reviewer's analysis was based on the completers.

The primary endpoint was Emax of Drug Liking VAS. On the average, the responses to each dose of eslicarbazepine acetate were significantly lower than those to any dose of alprazolam for Drug Liking VAS. The mean differences ranged from -22.44 to -12.39. In the comparison between eslicarbazepine acetate and placebo, on the average, the responses to three high doses of eslicarbazepine acetate (1600 mg, 2000 mg, and 2400 mg) were significantly higher than those to placebo, and there was no significant difference in responses between eslicarbazepine acetate 800 mg and placebo. There was apparent positive dose response for eslicarbazepine acetate. The comparison between each dose of alprazolam and placebo successfully validated the study.

Per CSS request, this reviewer also studied 4 secondary abuse potential measures: ARCI PCAG, Good Effects VAS, High VAS, and ARCI MBG. The study results showed that on the average, the responses to three high doses of eslicarbazepine acetate (1600 mg, 2000 mg, and 2400 mg) were significantly lower than those to both doses of alprazolam for all four measures in reviewer's secondary analysis. There was no significant mean (or median) difference between any dose of eslicarbazepine acetate and placebo for ARCI MBG. There was no significant mean (or median) difference between eslicarbazepine acetate 800 mg and placebo for ARCI PCAG. For other measures, on the average, responses to eslicarbazepine acetate 800 mg were significantly larger than those to placebo.

In summary, eslicarbazepine acetate is not euphoric, and has less liking, high, and good effects as well as sedative effects compared to alprazolam. However, high doses of eslicarbazepine acetate showed significant differences from placebo. Based on both the reviewer's primary and secondary analyses, this reviewer concludes that eslicarbazepine acetate has detectable abuse potential compared to placebo, and lower abuse potential than alprazolam.

2. Review Report on Study 093-153

2.1 Overview

2.1.1 Objectives of the study

Primary objective

The primary objective of this study was to evaluate the abuse potential of single doses of eslicarbazepine acetate compared to placebo and alprazolam in recreational CNS depressant users.

Secondary objectives

The secondary objective of this study was to evaluate the safety and tolerability of eslicarbazepine acetate in recreational CNS depressant users.

Reviewer's comment: This review report is only for the primary objective of the study.

2.1.2 Study design

The design was a single-dose, randomized, double-blind, and placebo- and active-controlled cross-over study with 7 Treatment Visits per subject.

The abuse potential of 4 doses of eslicarbazepine acetate (800 mg, 1600 mg, 2000 mg, and 2400 mg) was compared to that of placebo and 1.5 mg and 3.0 mg alprazolam (active control) in healthy recreational CNS depressant users. Subjects participated in a Screening Visit, one 4-day Qualification Phase, a Treatment Phase consisting of seven 3-day in-clinic Treatment Visits (each separated by a minimum 7-day washout period), and a safety Follow-up visit.

Within 28 days of a standard medical Screening, subjects attended a randomized, double-blind Qualification Phase in which they received either 2.0 mg alprazolam (Treatment Y) or matching placebo (Treatment X) in a cross-over manner, each separated by approximately 24 hours, to ensure that they could discriminate and show positive effects of alprazolam.

Following Qualification, it was planned that approximately 49 healthy female and male subjects aged 18 to 55 years (inclusive), who were recreational CNS depressant drug users and who had passed the pharmacologic Qualification, were randomized in the Treatment Phase. The treatments in the Treatment Phase were:

- Treatment A: 800 mg eslicarbazepine acetate
- Treatment B: 1600 mg eslicarbazepine acetate
- Treatment C: 2000 mg eslicarbazepine acetate
- Treatment D: 2400 mg eslicarbazepine acetate
- Treatment E: 1.5 mg alprazolam (active control)
- Treatment F: 3.0 mg alprazolam (active control)
- Treatment G: Placebo

Subjects were randomized to one of 14 treatment sequences according to a two 7 x 7 Williams square design. The study had a double-blind design such that the capsules and/or tablets received at each Treatment Visit were identical.

Treatment Visits were separated by a washout interval of at least 7 days. Subjects returned for the safety Follow-up Visit approximately 5 to 10 days following the last study drug administration.

2.1.3 Abuse potential measure and data collection times

The following pharmacodynamic assessments were administered to evaluate the subjective and objective effects of eslicarbazepine acetate.

Primary endpoint

Emax of Drug Liking VAS

Secondary endpoints

- Balance of effects:
 - Drug Liking VAS (Emin, and TA_AUE)
 - Overall Drug Liking VAS (Emax/Emin, end-of-day and next day scores)
 - Take Drug Again VAS (Emax, end-of-day and next day scores)
 - Subjective Drug Value (SDV; Emax, end-of-day and next day scores)
- Positive effects:
 - High VAS (Emax and TA_AUE)
 - Good Effects VAS (Emax and TA_AUE)
 - ARCI MBG (Emax and TA_AUE)
- Negative effects:
 - Bad Effects VAS (Emax and TA_AUE)
 - ARCI LSD (Emax and TA_AUE)
- Sedative effects:
 - ARCI PCAG scale (Emax and TA_AUE)
 - Alertness/Drowsiness VAS (Emin and TA_AUE)
 -
- Other drug effects:
 - Any Effects VAS (Emax and TA_AUE)
 - Drug Similarity VAS (score at 10-hours)
 - Dizziness VAS (Emax and TA_AUE)
 -
- Objective assessment of drug effects:
 - CRT
 - TRT (Emax and TA_AUE)
 - RRT (Emax and TA_AUE)
 - MRT (Emax and TA_AUE)
 -
- Percentage correct responses (Emin and TA_AUE)
 - DAT

- Mean percentage over road: percentage of time over the road (%; Emin and TA_AUE)
- TA_AUE)
- Mean response latency of correct responses (ms; Emax and TA_AUE)
- Number of false alarms (Emax and TA_AUE)
- Percentage of target hits (%; Emin and TA_AUE)
- HVLТ-R
 - Total recall (Emax and TA_AUE)
 - Delayed recall (Emax and TA_AUE)
 - Total number of errors (Emax and TA_AUE)
 - Retention (% retained; Emin and TA_AUE)
 - Recognition Discrimination Index (Emax and TA_AUE)

Reviewer's Comments: There were too many abuse potential measures in this study. The reviewer is wondering how a subject could respond to so many questions within even 1 hour, and how reliable the answers from the subjects to the questionnaires would be.

2.1.4 Number of subjects

Of the 53 subjects who were randomized to the Treatment Phase, 49 subjects (92.5%) were included in the PK Population and 44 subjects (83.0%) completed all treatments sessions, had no major protocol violations, and were included in the PD Population. Four subjects were excluded from the PK Population because they were withdrawn prior to receiving any eslicarbazepine acetate dose. Nine subjects were excluded from the PD Population because they were withdrawn prior to completing all treatment sessions of the study.

2.1.5 Statistical methodologies used in the Sponsor's analyses

The primary endpoint was analyzed using a mixed-effects model for a crossover study. The model included period, sequence, and treatment as fixed effects, as well as subject nested within sequence as a random effect, and baseline (pre-dose observation) as a covariate, where available. A first-order carryover effect was included in the model, but it was dropped if it was found to be non-significant at the 25% level. Statistical testing of the Abuse Potential Hypothesis was conducted by comparing the eslicarbazepine acetate treatments with placebo employing the Benjamini and Hochberg procedure (Benjamini-1995). Unadjusted *P* values were also provided. The interpretation of study results was based primarily on clinically meaningful differences, in conjunction with *P* values.

A second analysis to support the assessment of the Abuse Potential Hypothesis was conducted by comparing the eslicarbazepine acetate treatments to each validated dose of the active control (alprazolam: 1.5 mg and 3.0 mg; see Assay Sensitivity Hypothesis below). The endpoint for the supportive analysis was Emax of the Drug Liking VAS. The same statistical model and multiple testing procedures, as specified in the primary analysis, were employed in the supportive analysis. If both alprazolam doses were validated (ie, separated statistically from placebo treatment), then the supportive analysis was to be conducted twice employing 2 separate sets of pairwise comparisons.

An analysis to assess the Assay Sensitivity Hypothesis was conducted by comparing the active control (alprazolam: 1.5 mg and 3.0 mg) with placebo. The same statistical model as specified in

the primary analysis was employed; however, pairwise comparisons between the alprazolam doses and placebo were to be conducted without correction for multiple comparisons. If the 3.0 mg dose of alprazolam had statistically significantly ($\alpha = 0.05$; two-tailed) higher abuse potential than placebo, then the study was to be considered valid for the determination of abuse potential in eslicarbazepine acetate. The endpoint for analysis of the Assay Sensitivity Hypothesis was Emax of the Drug Liking VAS.

All secondary endpoints were also evaluated using a similar mixed-effect model (without correction for multiplicity) for comparisons of each eslicarbazepine acetate dose compared to placebo and each alprazolam dose.

For all subjective dependent measures, Emax was used for positive scores (liking) and Emin was used for negative scores (disliking). TA_AUE values were calculated and reported as supportive endpoints. Data was summarized graphically, where appropriate. Dose response of the study drugs were also examined graphically. The time course scores for all PD measures were listed for all randomized subjects.

For each of the primary and secondary measures, the scores at each time point and derived parameters were summarized using descriptive statistics (eg, N, mean, SD, median, minimum, maximum, and 95% confidence intervals [CI]) by treatment.

All analyses were investigated against the statistical assumptions implicit within that analysis; failure of those statistical assumptions (for example, distributional violations) resulted in a changed analysis to account for the true apparent features of the data (eg, rank transformation of Emax of the Drug Liking VAS in the primary analysis).

2.1.5 Sponsor's Summary and Conclusions

Summary

- Drug Liking Emax (primary endpoint) for both doses of alprazolam was significantly greater than placebo, thereby confirming the validity of the study. The 3 highest eslicarbazepine acetate doses were significantly different from placebo, and all eslicarbazepine acetate doses showed significantly lower Drug Liking VAS Emax values compared to 1.5 mg and 3.0 mg alprazolam. Results were similar with secondary Drug Liking VAS endpoints (Emin and TA_AUE); however, treatment differences were more modest.
- On secondary balance of effects measures (Overall Drug Liking VAS, Take Drug Again VAS, and SDV), eslicarbazepine acetate showed significant differences from placebo primarily at the higher doses (eg, 2000 mg and/or 2400 mg) and only on some endpoints. In contrast, both alprazolam doses showed significantly greater effects compared to placebo on all endpoints. All eslicarbazepine acetate doses showed significantly lower effects compared to both alprazolam doses on the majority of endpoints. In general, there was a plateau in the dose-response at the higher eslicarbazepine acetate doses.
- On secondary positive effects measures (Good Effects VAS, High VAS and ARCI MBG), significant differences were observed between eslicarbazepine acetate doses and placebo on most endpoints, particularly at the higher doses, although on the ARCI MBG

(“euphoria”) scale, only the 2400 mg dose was significantly different from placebo. Both alprazolam doses were significantly different from placebo on all endpoints, while all eslicarbazepine acetate doses showed significantly lower effects compared to both alprazolam doses.

- Treatment effects on the negative effects measures (Bad Effects VAS and ARCI LSD) were more modest, although significant differences were observed on some endpoints between 2000 mg and 2400 mg eslicarbazepine acetate and placebo, as well as between alprazolam and placebo. Negative effects were generally greater with alprazolam (particularly at 3.0 mg) relative to eslicarbazepine acetate.
- The pattern of effects on sedative and other measures was similar to positive effects measures; eslicarbazepine acetate was generally different from placebo but showed significantly lower effects compared to alprazolam. On Drug Similarity VAS, while alprazolam was identified strongly as a benzodiazepine, and to a lesser extent codeine/morphine and other drugs, responses with eslicarbazepine acetate were more modest.
- Alprazolam was associated with significant, dose-dependent impairment on the majority of cognitive and psychomotor endpoints (CRT, DAT, and HVL-T-R), while cognitive effects of eslicarbazepine acetate were more modest and generally significantly lower than those of alprazolam.
- Dose-effect relationships were relatively shallow for both active treatments. Eslicarbazepine acetate plasma concentrations were correlated with mean Drug Liking VAS scores; the slope of the relationship was very shallow.
- Mean eslicarbazepine acetate C_{max} and AUC_{0-last} increased with increasing eslicarbazepine acetate dose. Median $t_{1/2}$ ranged from 13 to 17 hours and median t_{max} ranged between approximately 1 to 3 hours post-dose.
- Overall, eslicarbazepine acetate showed statistically significant subjective effects compared to placebo on most endpoints; however, the magnitude of the effects was minimal and significantly lower than that of alprazolam on the primary endpoint and all key secondary endpoints.

Conclusion

The statistically significant differences between eslicarbazepine acetate and placebo on the primary (and most secondary endpoints) indicate that it has detectable subjective effects and showed some drug 'liking' in recreational CNS depressant users at supratherapeutic doses; however, the magnitude of this effect was minimal and unlikely to be clinically relevant. The significant effects of alprazolam compared to placebo on Drug Liking VAS E_{max} and the majority of secondary endpoints demonstrate the validity of the study and sensitivity of the measures for detecting abuse-related effects, as well as cognitive and psychomotor impairment.

This study also demonstrated that single doses of eslicarbazepine acetate had less abuse potential than alprazolam in recreational depressant users. Eslicarbazepine acetate effects were significantly lower than alprazolam on the primary endpoint (Drug Liking VAS E_{max}) and the majority of secondary endpoints.

2.2 Data Location

The analysis dataset is located at

<\\cdsesub1\EVSPROD\NDA022416\0053\m5\53-clin-stud-rep\534-rep-human-pd-stud\5341-healthy-subj-pd-stud-rep\093-153\093-153-synop.pdf>

2.3 Reviewer's Assessment

In the reviewer's report P, A1.5, A3, E800, E1600, E2000, and E2400 denote placebo, alprazolam 1.5 mg and 3 mg, eslicarbazepine acetate 800 mg, 1600 mg 2000 mg and 2400, respectively.

2.3.1 Missing data issue

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

Figure 1 shows the individual time course response profiles for A3 for Drug Liking VAS. The orange line separates the responses by gender. The subjects above the orange line are females, and the subjects below the orange line are males. Colors blue, white, and red denote dislike, neutral and like, respectively. The grey color indicates missing data. From this figure, one may see that for A3 27.2% (12/44), 29.5% (13/44), 13.5% (6/44) and 9.1% (4/44) of subjects have missing data at hours 1, 1.5, 2, and 3, respectively. The missing data situation for A3 is similar to what has been observed in another human drug abuse potential study in a past NDA for recreational polydrug users. Figure 2 showed less missing data for A1.5 than observed for A3.

Figure 3 is the individual time course response profiles for E2400 for Drug Liking VAS. From this graph, one may see that only three subjects have missing data at hour 1.5.

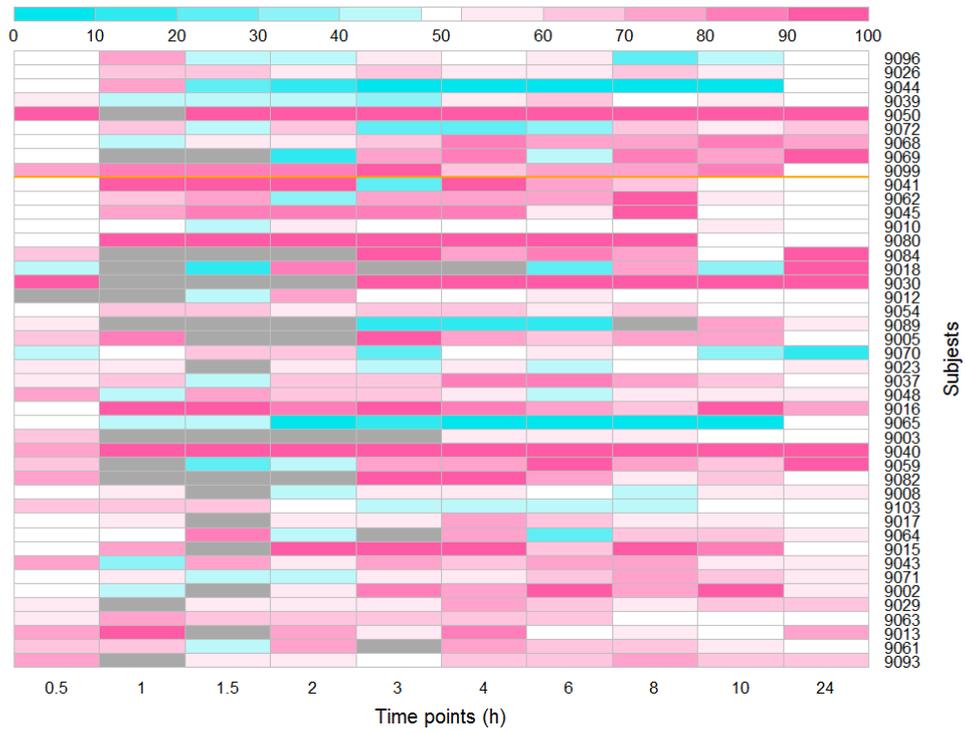


Figure 1: Individual time course response profiles for Drug Liking VAS (A3)

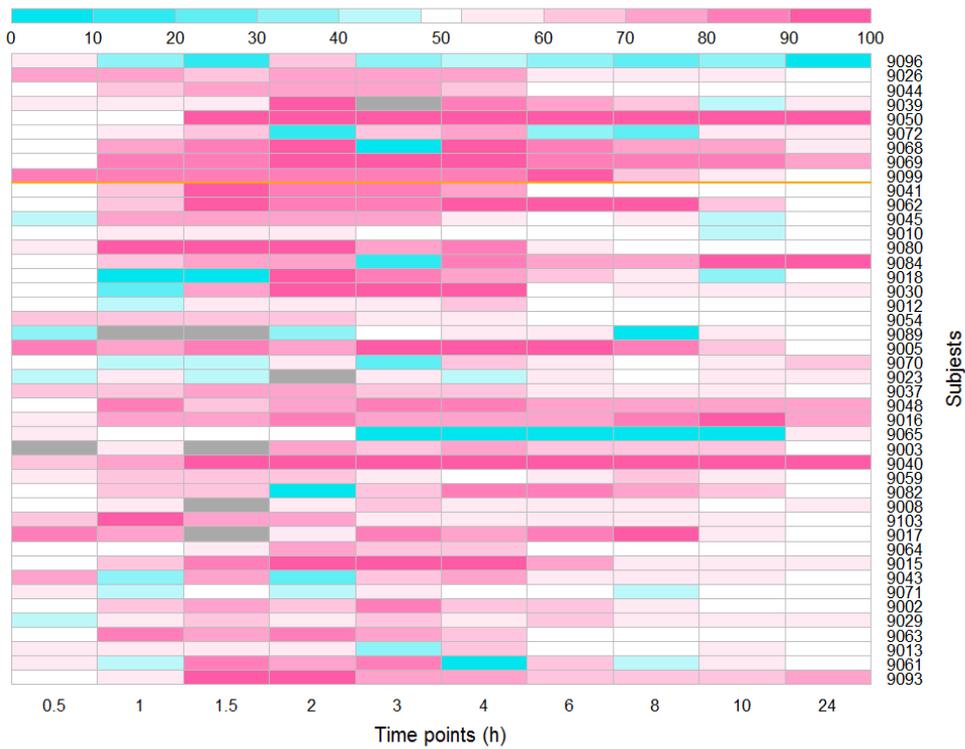


Figure 2: Individual time course response profiles for Drug Liking VAS (A1.5)

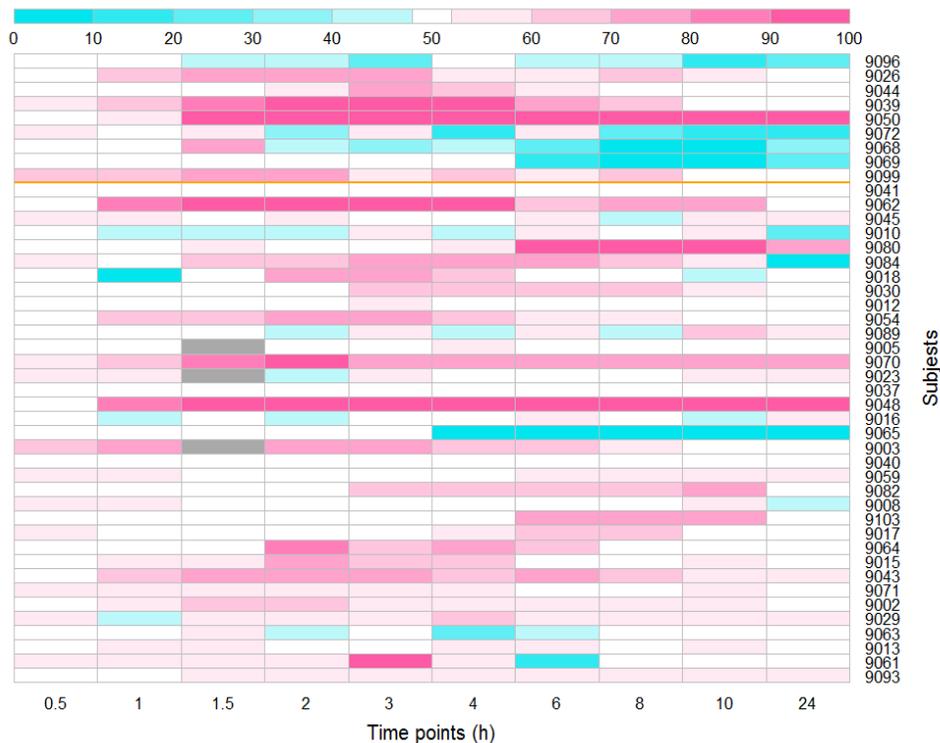


Figure 3: Individual time course response profiles for Drug Liking VAS (E2400)

A request for explanation of the missing data along with other questions from the CSS was sent to the sponsor on April 25, 2013. The sponsor responded on May 1, 2013 that

All subjects with missing data points in the Alprazolam (ALZ) 3 mg treatment period had somnolence sufficient to prevent them from providing an accurate assessment (i.e. fell asleep) and were not easily awakened. It is standard practice at the site conducting the study to attempt to arouse the subject at least once. If the subject is extremely sedated and is not easily aroused, the investigator must assess the subject in order for the collection of PD data to be skipped.

The detailed explanation and discussion of the missing data issue by the Sponsor can be found at

<http://cdsesub1\evsprod\nda022416\0083\m1\us\111-info-amend\rsp-to-2013-04-25-div-info-rqst.pdf>

Figure 4 shows the individual time course response profiles for A2 in the Qualification Phase for the subjects selected for the Treatment Phase. Compared to [Figure 1](#), less blue color and more red color are on Figure 4. It means that these subjects liked A2 more than A3. Only one subject has missing data at hour 0.5 for A2, which may not be due to AE from the drug effect.

Because this was a crossover study and the primary endpoint was Emax, the missing data were not imputed in either the Sponsor's analysis or the reviewer's analysis. However, both the Sponsor and CSS may need to consider if A3 is a proper dose for the active control for such studies. If subjects who experienced AEs belong to a special subgroup of the study population,

the sample size for future studies needs to be increased, and stratification may need to be considered in the study design stage.

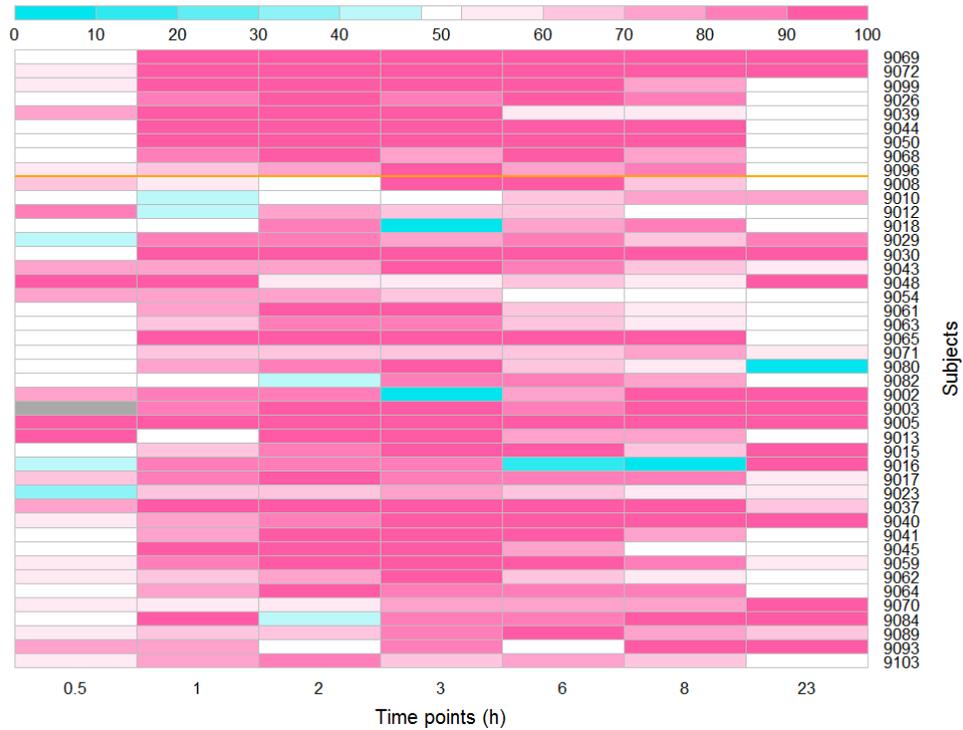


Figure 4: Individual time course response profiles for Drug Liking VAS (A2 in Qualification Phase)

2.3.2 Primary Analysis

2.3.2.1 Descriptive Statistics

Table 1 summarizes the mean, standard error, minimum, the first quartile (Q_1), median, the third quartile (Q_3), and maximum for seven treatments in the study and for the treatment differences between eslicarbazepine acetate and alprazolam or placebo for Emax of Drug Liking VAS.

Table 1: Summary statistics for Emax of Drug Liking VAS (N=44)

TRT or Comparison	Mean	StdErr	Min	Q1	Med	Q3	Max
A1.5	81.05	2.41	51	68	82.5	98.25	100
A3	80.55	2.44	50	69	77.5	98.75	100
E800	58.18	1.67	50	51	51.5	63.75	100
E1600	64.43	2.43	50	51	61	72.25	100
E2000	66.89	2.34	50	52.5	63	75.75	100
E2400	68.23	2.59	50	51	68	78.75	100
P	55.20	1.73	50	51	51	52.5	100
E800-A1.5	-22.86	2.82	-50	-40	-22.5	-10.25	9
E800-A3	-22.36	3.09	-50	-39.75	-23.5	-4	34
E800-P	2.98	2.33	-49	0	0	8	49
E1600-A1.5	-16.61	2.88	-49	-29.75	-13.5	-2.5	32
E1600-A3	-16.11	3.07	-50	-31.5	-15	0	25
E1600-P	9.23	2.09	-15	0	4	16	50
E2000-A1.5	-14.16	2.37	-42	-25	-12.5	-1.25	18
E2000-A3	-13.66	2.76	-50	-26.75	-12.5	0	25
E2000-P	11.68	2.31	-15	0	8	22.5	50
E2400-A1_5	-12.82	2.73	-50	-26	-13	0	24
E2400-A3	-12.32	3.44	-50	-29.25	-12.5	5.5	34
E2400-P	13.02	2.82	-30	0	9	26	50

Table 1 shows that the third quartiles of A1.5 and A3 are over 98. It means that even for a schedule IV drug, alprazolam, the Emax of Drug Liking VAS could be extremely large in approximately 25% of subjects. One may notice that the means and medians of the differences between eslicarbazepine acetate and alprazolam are all negative. The means and medians of the differences between eslicarbazepine acetate and placebo are all positive except the zero median difference between E800 and P.

Figure 5 provides the boxplots of five treatments as well as boxplots for the differences between eslicarbazepine acetate and alprazolam, and between eslicarbazepine acetate and placebo for Drug Liking VAS. The line in each box denotes the median and the circle in each box is for the mean. Because of the rules of the sorting process for characters in SAS, the boxplots related to E800 appear after those related to E2400 on the graph. Compared to alprazolam, majority subjects (in most cases approximately 75% of subjects) have lower Emax for eslicarbazepine acetate. For the comparison between each dose of eslicarbazepine acetate and placebo, the majority of the differences are greater than zero.

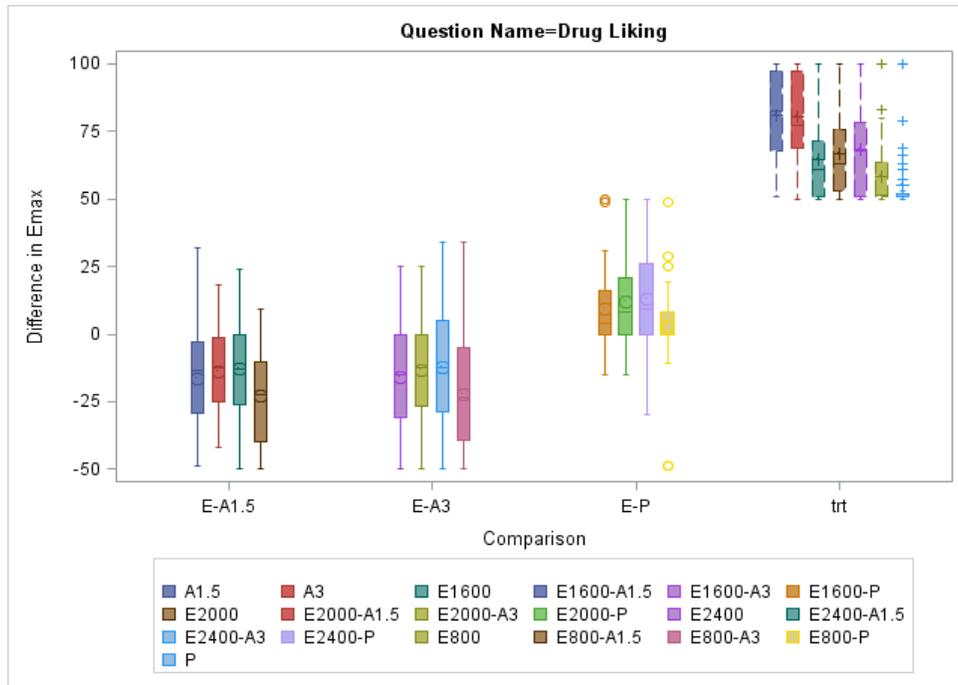


Figure 5: Boxplots for seven treatments and the differences between eslicarbazepine acetate and alprazolam, and between eslicarbazepine acetate and placebo for Drug Liking VAS (N=44)

Figure 6 plots the mean and median dose response curves for eslicarbazepine acetate and alprazolam for Drug Liking VAS. The figures show a positive dose response for eslicarbazepine acetate. The means of A1.5 and A3 are similar. The median of A3 is lower than that of A1.5. This scenario may be due to the AEs experienced by some subjects from A3.

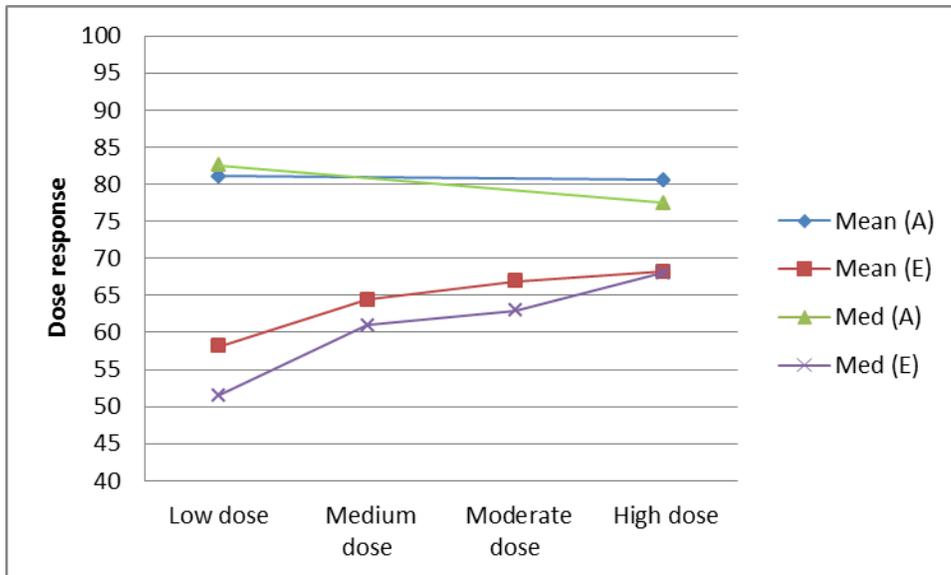


Figure 6: Mean and median dose response curves for Drug Liking VAS (N=44)

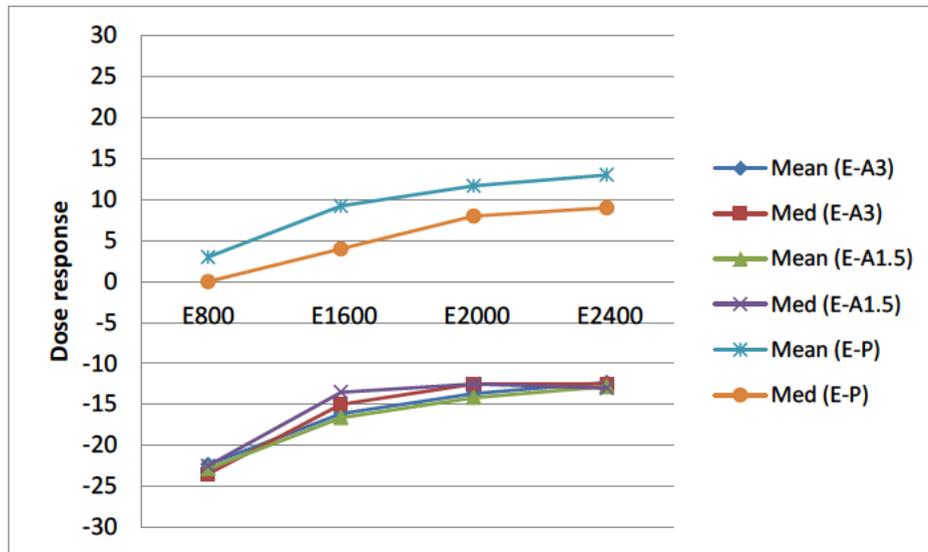


Figure 7: Dose response curves of eslicarbazepine acetate in difference from alprazolam and placebo for Drug Liking VAS (n=44)

Because this is a crossover study, the dose response curves for the test drug relative to the active control and placebo may be also useful. Figure 7 plots the mean and median dose response curves for eslicarbazepine acetate in difference from alprazolam and placebo for Drug Liking VAS. Again, a positive dose response of eslicarbazepine acetate is observed.

Figure 8 is the mean time course profiles for Drug Liking VAS. Because of the missing data issue, the means at early time points were calculated using data from subjects who had responses at these time points. Four profiles from eslicarbazepine acetate are under those of alprazolam. The profile for E800 is very similar to that of placebo. The A3 has three peaks at hours 2, 4, and 8, and A1.5 has two peaks at hours 2 and 5. This may be due to missing data in the calculation for the means at early time points. The peak mean response of A1.5 is larger than that of A3.

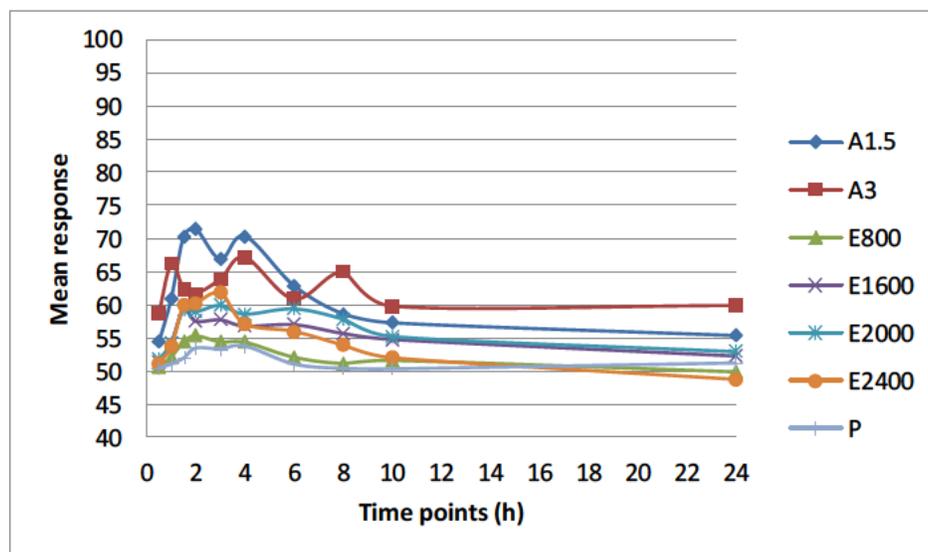


Figure 8: Mean time course profiles for Drug Liking VAS (N=44)

Figure 9 is the heat map display for Emax of Drug Liking VAS by treatment. From this graph, one may notice that some subjects have high placebo responses. The responses to E800 and placebo are very similar. Overall, more subjects highly liked alprazolam compared to eslicarbazepine acetate.

The light pink in the category [51, 60] occurs in many subjects for placebo and E800. One may notice that in [Table 1](#), the medians of Emax of Drug Liking VAS are 51.5 and 51 for E800 and P respectively. This reviewer is wondering if the neutral score is originally set up at 51 instead of 50 on the computer screens for these subjects. This reviewer does not believe that subjects would be able to move the middle bar on the bipolar visual analog scale for only one unit away from 50, because the middle bar itself looks more than one unit wide.

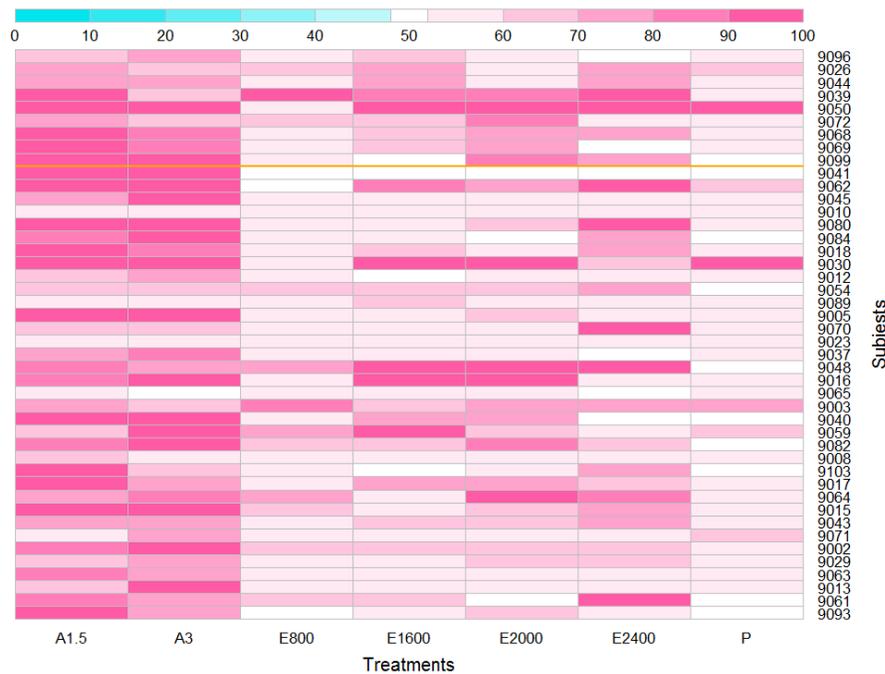


Figure 9: Heat map for Emax of Drug Liking VAS by treatment

2.3.2.2 Statistical Testing

The statistical model used in the reviewer’s analysis was the mixed-effects model with period, sequence, and treatment as fixed effects, and subject as a random effect. The reviewer checked assumptions in the model for the equal variances and the normality. The normal assumption was not violated for Drug Liking VAS. However, the assumption of equal variances was not satisfied. The SAS proc mixed procedure can adjust the unequal variances using Tukey-Kramer’s method.

Table 2 shows the analysis results. The least square mean and 95% confidence interval for the mean of each treatment are shown in the third row of Table 2. Rows 4-6 show the mean differences and p-values for the comparisons between eslicarbazepine acetate and each dose of alprazolam, and between eslicarbazepine acetate and placebo.

Table 2: Statistical analysis results for Drug Liking VAS

Measure	Treatment	E800	E1600	E2000	E2400	A1.5	A3	P
	N	44	44	44	44	44	44	44
Drug Liking VAS	LS mean	58.12	64.26	66.84	68.17	81.02	80.56	55
	95% CI	(53.72, 62.51)	(59.87, 68.66)	(62.44, 71.23)	(63.78, 72.57)	(76.62, 85.42)	(76.17, 84.96)	(50.61, 59.40)
	Diff vs A1.5/pval	-22.9 / <.0001	-16.76 / <.0001	-14.18 / <.0001	-12.85 / <.0001			
	Diff vs A3/pval	-22.44 / <.0001	-16.30 / <.0001	-13.72 / <.0001	-12.39 / <.0001			
	Diff vs P/pval	3.11 / 0.9070	9.26 / 0.0112	11.83 / 0.0003	13.17 / <.0001			

Note: pval denotes p-value. All p-values were from the two-sided t test, and adjusted by Tukey -Kramer’s method for unequal variances.

The primary analysis show that

On the average, the responses to each dose of eslicarbazepine acetate were significantly lower than those to any dose of alprazolam for Drug Liking VAS. The mean differences were ranged from -22.44 to -12.39. In the comparison between eslicarbazepine acetate and placebo, on the average, responses to three high doses (E1600, E2000, and E2400) were significantly higher than those to placebo. There was no significant difference between E800 and P.

2.3.3 Secondary Analysis

Per the CSS reviewer Dr. Alicja Lerner’s request, the reviewer’s secondary analysis included abuse potential measures: ARCI MBG, ARCI PCAG, Good Effects VAS, and High VAS.

The same methodologies as the primary analysis were used in the secondary analysis. Among four secondary measures, the normal assumption of the model is satisfied for ARCI PCAG and High VAS.

Table 3 summarizes the test results for ARCI MBG and High VAS. The adjusted p-values are from Tukey-Kramer’s method for unequal variances.

Table 3: Statistical analysis results for ARCI MBG and High VAS

Measure	Treatment	E800	E1600	E2000	E2400	A1.5	A3	P
	N	44	44	44	44	44	44	44
High VAS	LS mean	38.09	47.46	62.03	58.19	82.68	89.31	19.96
	95% CI	(28.79, 47.39)	(38.16, 56.76)	(52.74, 71.33)	(48.89, 67.48)	(73.38, 91.98)	(80.01, 98.60)	(10.66, 29.26)
	Diff vs A1.5/pval	-44.59 / <.0001	-35.22 / <.0001	-20.64 / 0.0070	-24.49 / 0.0005			
	Diff vs A3/pval	-51.22 / <.0001	-41.85 / <.0001	-27.27 / <.0001	-31.12 / <.0001			
	Diff vs P/pval	18.13 / 0.0292	27.50 / <.0001	42.07 / <.0001	38.23 / <.0001			
ARCI PCAG	LS mean	3.83	4.61	5.47	5.53	8.22	8.83	2.88
	95% CI	(2.79, 4.88)	(3.57, 5.66)	(4.43, 6.51)	(4.48, 6.57)	(7.17, 9.26)	(7.79, 9.88)	(1.84, 3.93)
	Diff vs A1.5/pval	-4.39 / <.0001	-3.6 / <.0001	-2.75 / <.0001	-2.69 / <.0001			
	Diff vs A3/pval	-5.00 / <.0001	-4.22 / <.0001	-3.36 / <.0001	-3.3 / <.0001			
	Diff vs P/pval	0.95 / 0.6258	1.73 / 0.0369	2.59 / 0.0001	2.65 / <.0001			

Note: pval denotes p-value. All p-values were from two-sided t tests, and adjusted by Tukey -Kramer’s method for unequal variances.

Table 3 shows that eslicarbazepine acetate had lower least square means than alprazolam for both ARCI PCAG and High VAS, and the differences were statistically significant. For the comparison between eslicarbazepine acetate and placebo, except E800 versus P (p-value=0.6258)

for ARCI PCAG, the mean difference between each dose of eslicarbazepine acetate and placebo was statistically significantly greater than 0.

Tables 4-7 are summary statistics for 7 treatments in the study and for the treatment differences between eslicarbazepine acetate and alprazolam (or placebo) for Emax of ARCI MBG, ARCI PCAG, Good Effects VAS and High VAS. The mean time course profiles for these measures are presented in [5.1 Appendix I](#). Excluding ARCI PCAG and High VAS, if the statistical test for the comparison between two treatments is significant based on the paired t test, the mean is highlighted in red in these tables; if the statistical test for the comparison between two treatments is significant based on the Sign test or the Wilcoxon Signed-Rank test, the median is highlighted in blue. The p-values of the tests are provided in [5.2 Appendix II](#).

Table 4: Summary statistics and testing results for Emax of ARCI MBG

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
A1.5	44	6.89	0.86	0	2	6	13	16
A3	44	8.23	0.75	0	4	8.5	12.75	16
E800	44	1.80	0.46	-1	0	0	2	12
E1600	44	1.84	0.58	-2	0	0	1	14
E2000	44	1.57	0.47	-4	0	0	1.75	14
E2400	44	2.30	0.58	-1	0	1	3	15
P	44	1.20	0.44	-1	0	0	1	14
E800-A1.5	44	-5.09	0.89	-16	-10.75	-3	-1	6
E800-A3	44	-6.43	0.80	-16	-11.75	-6.5	-2	1
E800-P	44	0.59	0.64	-13	-1	0	2	12
E1600-A1.5	44	-5.05	0.91	-16	-10.75	-3.5	0	6
E1600-A3	44	-6.39	0.91	-16	-12	-7	-2	9
E1600-P	44	0.64	0.49	-11	0	0	1	10
E2000-A1.5	44	-5.32	0.88	-16	-10.75	-3	-0.25	3
E2000-A3	44	-6.66	0.84	-16	-12	-6.5	-2	6
E2000-P	44	0.36	0.39	-9	0	0	1	7
E2400-A1.5	44	-4.59	0.85	-15	-9.5	-2.5	-0.25	5
E2400-A3	44	-5.93	0.89	-16	-11.75	-6.5	-2	7
E2400-P	44	1.09	0.69	-12	-0.75	0	2.5	15

Table 5: Summary statistics for Emax of ARCI PCAG

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
A1.5	44	8.2955	0.4259	1	6.25	9	10	13
A3	44	8.9091	0.3514	4	8	9	11	14
E800	44	3.8864	0.5782	-1	0	3	6	14
E1600	44	4.6818	0.5535	0	1	4.5	8	12
E2000	44	5.5682	0.6155	-4	3	5	9	14
E2400	44	5.6364	0.5753	0	2	6	9	14
P	44	3	0.5372	-1	0	1.5	5	13
E800-A1.5	44	-4.409	0.5473	-11	-7.75	-4.5	-1	1
E800-A3	44	-5.023	0.5937	-12	-8.75	-5	-2	2
E800-P	44	0.8864	0.648	-5	-2	0	3	14
E1600-A1.5	44	-3.614	0.4166	-9	-5	-4	-2	6
E1600-A3	44	-4.227	0.5004	-11	-6	-4.5	-2	3
E1600-P	44	1.6818	0.7039	-9	-1	1	5	12
E2000-A1.5	44	-2.727	0.4793	-14	-4	-2	-1	2
E2000-A3	44	-3.341	0.5733	-15	-6	-3	0	4
E2000-P	44	2.5682	0.6761	-7	-0.75	2	5	14
E2400-A1.5	44	-2.659	0.4876	-11	-4	-2.5	-1	3
E2400-A3	44	-3.273	0.5316	-11	-6	-3	-1	3
E2400-P	44	2.6364	0.7124	-6	-0.75	2	5	14

Note: The statistical test results are shown in [Table 3](#).

Table 6: Summary statistics and testing results for Emax of Good Effects VAS

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
A1.5	44	86.932	2.709	7	79.25	92.5	100	100
A3	44	85.591	2.5004	49	72.25	90.5	100	100
E800	44	46.75	4.587	0	9.75	54	66.75	100
E1600	44	49.273	5.6455	0	4.25	59.5	77.25	100
E2000	44	64.182	4.3286	0	51	69	82.75	100
E2400	44	61.886	5.1461	0	50	68.5	94.25	100
P	44	21.636	4.6656	0	0	1	51	100
E800-A1.5	44	-40.18	5.159	-100	-56	-37	-19.5	64
E800-A3	44	-38.84	5.1314	-100	-56.75	-35.5	-14.5	17
E800-P	44	25.114	5.9278	-100	0	18	57.25	100
E1600-A1.5	44	-37.66	5.21	-100	-73.25	-26.5	-8	7
E1600-A3	44	-36.32	5.6707	-100	-71.5	-30	-3	49
E1600-P	44	27.636	5.9918	-56	0	15.5	61.25	100
E2000-A1.5	44	-22.75	3.5683	-90	-34.25	-22	-1.5	14
E2000-A3	44	-21.41	4.1986	-100	-38.25	-21	0	51
E2000-P	44	42.545	5.8521	-56	13	49.5	69.75	100
E2400-A1.5	44	-25.05	4.8713	-100	-37	-19	0	50
E2400-A3	44	-23.7	5.0761	-100	-48.5	-19	0	49
E2400-P	44	40.25	6.564	-65	1	48.5	74.75	100

Table 7: Summary statistics for Emax of High VAS

TRT/ Comparison	N	Mean	StdErr	Min	Q1	Med	Q3	Max
A1.5	44	82.523	3.6002	0	69.25	92.5	100	100
A3	44	89.068	2.3919	32	80.5	100	100	100
E800	44	37.568	5.3243	0	0	43.5	65.25	100
E1600	44	47.045	5.9121	0	2	59	76.5	100
E2000	44	61.909	4.5486	0	50.25	65.5	86.25	100
E2400	44	58.318	5.2151	0	44.25	67	84.5	100
P	44	20.182	4.5712	0	0	1	51	100
E800-A1.5	44	-44.95	6.6669	-100	-89.75	-45	-15.75	90
E800-A3	44	-51.5	6.0084	-100	-90.75	-47.5	-25.5	39
E800-P	44	17.386	6.5045	-79	0	3	54.75	100
E1600-A1.5	44	-35.48	5.4338	-100	-70	-26.5	-1.25	14
E1600-A3	44	-42.02	5.7333	-100	-79.75	-31	-0.25	21
E1600-P	44	26.864	6.713	-67	0	8	66.75	100
E2000-A1.5	44	-20.61	3.5052	-92	-38.5	-13.5	0	12
E2000-A3	44	-27.16	4.6174	-100	-47.25	-26.5	0	40
E2000-P	44	41.727	5.7702	-59	7.25	45.5	72.75	100
E2400-A1.5	44	-24.2	5.192	-100	-39.5	-25	-0.25	90
E2400-A3	44	-30.75	5.3231	-100	-53.25	-26	-1.25	39
E2400-P	44	38.136	7.041	-100	6.25	46.5	75	100

Note: The statistical test results are shown in [Table 3](#).

Tables 4 and 6 show that

For ARCI MBG, on the average, the responses to each dose of eslicarbazepine acetate were significantly lower than those to both doses of alprazolam, and there was no significance difference between eslicarbazepine acetate and placebo.

For Good Effects VAS, on the average, the responses to each dose of eslicarbazepine acetate were significantly lower than those to both doses of alprazolam, and significantly greater than those to placebo.

3. Conclusion

The results from both the reviewer's primary and secondary analyses are summarized in Table 8.

Table 8: Summary of the results from significance tests for the abuse potential measures considered in this review

Comarison	Drug Liking VAS	High VAS	ARCI PCAG	Good Effects VAS	ARCI MBG
A1.5-P	S (>)	S (>)	S (>)	S (>)	S (>)
A3-P	S (>)	S (>)	S (>)	S (>)	S (>)
E800-A1.5	S (<)	S (<)	S (<)	S (<)	S (<)
E800-A3	S (<)	S (<)	S (<)	S (<)	S (<)
E800-P	NS	S (>)	NS	S (>)	NS
E1600-A1.5	S (<)	S (<)	S (<)	S (<)	S (<)
E1600-A3	S (<)	S (<)	S (<)	S (<)	S (<)
E1600-P	S (>)	S (>)	S (>)	S (>)	NS
E2000-A1.5	S (<)	S (<)	S (<)	S (<)	S (<)
E2000-A3	S (<)	S (<)	S (<)	S (<)	S (<)
E2000-P	S (>)	S (>)	S (>)	S (>)	NS
E2400-A1.5	S (<)	S (<)	S (<)	S (<)	S (<)
E2400-A3	S (<)	S (<)	S (<)	S (<)	S (<)
E2400-P	S (>)	S (>)	S (>)	S (>)	NS

Note: The sign (>) shows that in comparison of A versus B, on the average, A was greater than B. The (<) sign denotes that on the average, A was smaller than B. S and NS note significant difference and nonsignificant difference, respectively. Blue is for S (>), and red is for S (<).

This study shows that

- The comparison between each dose of alprazolam and placebo successfully validated the study.
- There was no significant mean (or median) difference between any dose of eslicarbazepine acetate and placebo for ARCI MBG.
- There was no significant mean (or median) difference between eslicarbazepine acetate 800 mg and placebo for Drug Liking VAS and ARCI PCAG. For other measures, on the average, responses to eslicarbazepine acetate 800 mg were significantly larger than those to the placebo.
- On the average, the responses to three high doses of eslicarbazepine acetate (E1600, E2000 and E2400) were significantly lower than those to both doses of alprazolam for all five measures in the reviewer's primary and secondary analyses.
- There was no mean dose-response for alprazolam in this study. The median dose-response was negative due to AEs of alprazolam 3 mg.
- There is positive mean (and median) dose-response for eslicarbazepine acetate.

In summary, eslicarbazepine acetate is not euphoric, and has less liking, high, and good effects as well as sedative effects compared to alprazolam. However, three high doses of eslicarbazepine

acetate showed significant differences from placebo. Based on both the primary and secondary analyses, this reviewer concludes that eslicarbazepine acetate has detectable abuse potential compared to placebo, and lower abuse potential than alprazolam.

4. Reference

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

5. Appendices

5.1 Appendix I

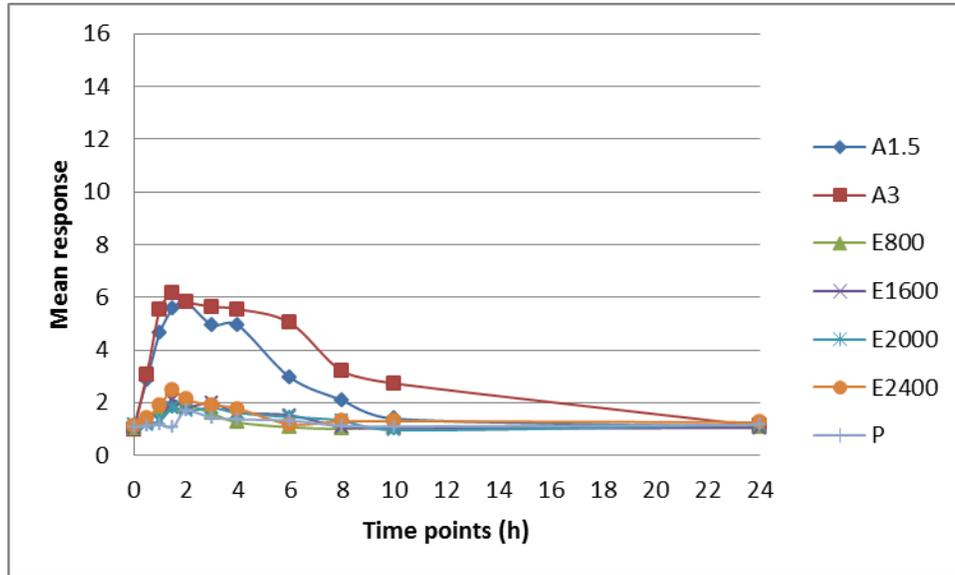


Figure 10: Mean time course profiles for ARCI MBG

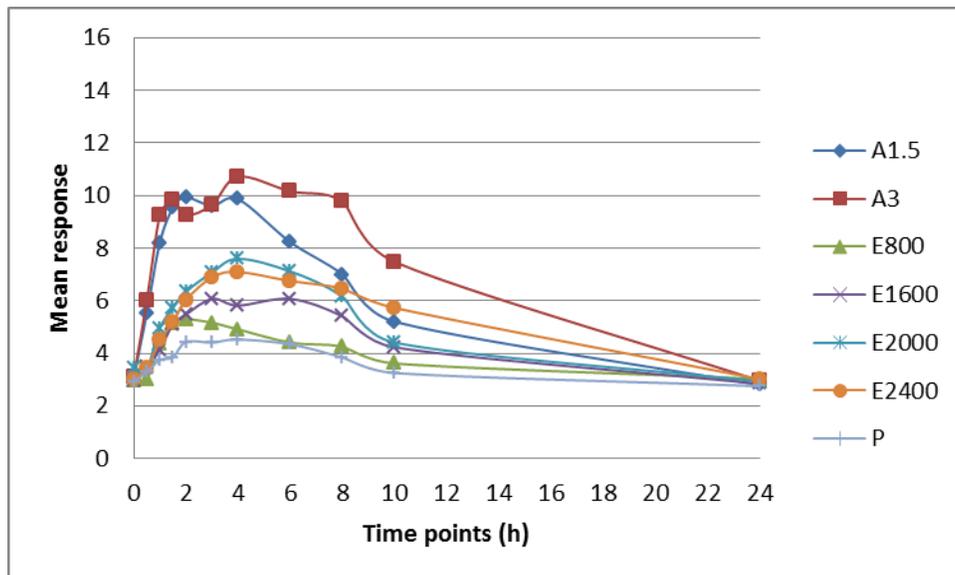


Figure 11: Mean time course profiles for ARCI PCAG

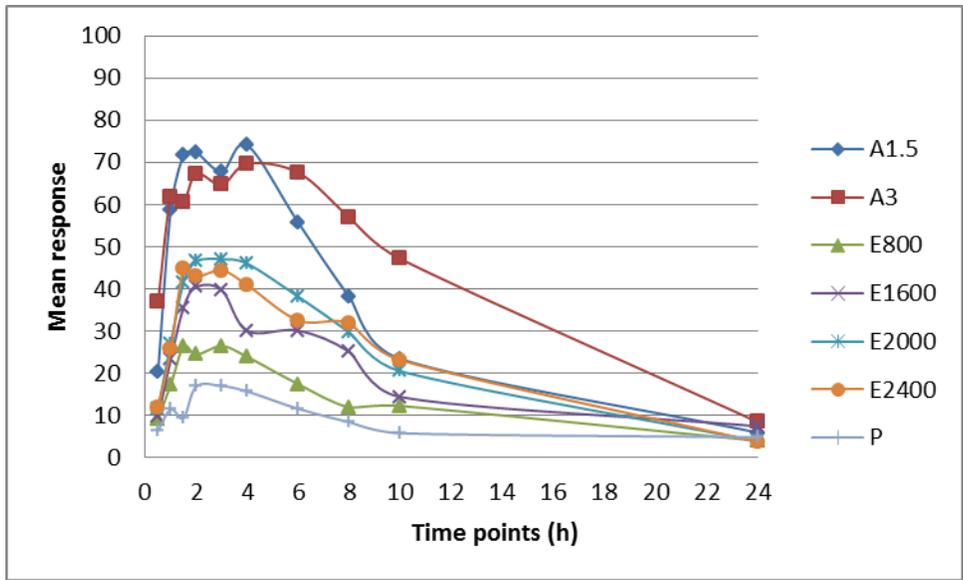


Figure 12: Mean time course profiles for Good Effects VAS

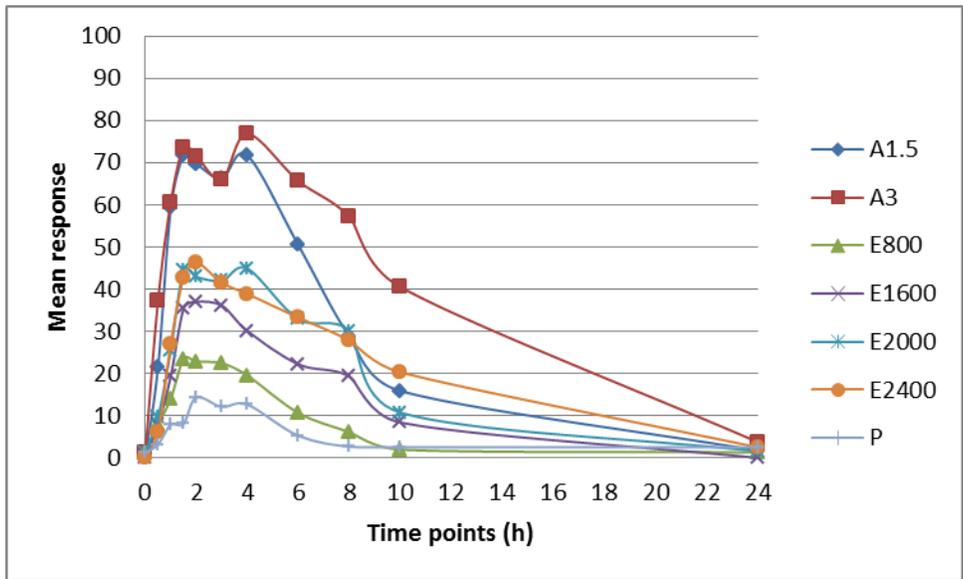


Figure 13: Mean time course profiles for High VAS

5.2 Appendix II

Table 9: P-values of the significance tests for Emax of ARCI MBG (n=44)

Comparison	t-test	Sign-test	Signed-Rank test	W-test
E800-A1.5	<.0001	<.0001	<.0001	0.0029
E800-A3	<.0001	<0.0001	<.0001	0.0091
E800-P	0.3621	0.7011	0.2102	<.0001
E1600-A1.5	<.0001	<.0001	<.0001	0.0114
E1600-A3	<.0001	<.0001	<.0001	0.2505
E1600-P	0.1980	0.5413	0.2648	<0.0001
E2000-A1.5	<.0001	<.0001	<.0001	0.0004
E2000-A3	<.0001	<.0001	<.0001	0.1440
E2000-P	0.3550	0.1686	0.0995	<.0001
E2400-A1.5	<.0001	<.0001	<.0001	0.0003
E2400-A3	<.0001	<0.0001	<.0001	0.0625
E2400-P	0.1214	0.4421	0.0958	<.0001

Note: The normality of the model is not satisfied for ARCI MBG. The red p-value indicates that the W-test was significant for the distribution of the paired differences, and the green p-value indicates that the test used in the evaluation was according to the assumption of the test. For example, In the case of E1600-A3, the W test was not significant. It means that the normal assumption for the distribution of differences in responses between E1600 and A3 was not violated. Thus, the t-test was used for the comparison, and resulted in a p-value <.0001 (two-sided).

Table 10: P-values of the significance tests for Emax of Good Effects (n=44)

Comparison	t-test	Sign-test	Signed-Rank test	W-test
E800-A1.5	<.0001	<.0001	<.0001	0.0554
E800-A3	<.0001	<.0001	<.0001	0.0127
E800-P	0.0001	0.0003	<.0001	0.0030
E1600-A1.5	<.0001	<.0001	<.0001	0.0006
E1600-A3	<.0001	<.0001	<.0001	0.0549
E1600-P	<.0001	0.0005	<.0001	0.0265
E2000-A1.5	<.0001	<.0001	<.0001	0.0282
E2000-A3	<.0001	0.0001	<.0001	0.3303
E2000-P	<.0001	<.0001	<.0001	0.0876
E2400-A1.5	<.0001	0.0001	<.0001	0.0013
E2400-A3	<.0001	0.0005	<.0001	0.0899
E2400-P	<.0001	0.0000	<.0001	0.0192

See the note in Table 9.

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

/s/

LING CHEN
05/31/2013

STELLA G MACHADO
05/31/2013



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

CLINICAL STUDIES

NDA/Serial Number: 22,416

Drug Name: Eslicarbazepine acetate

Indication(s): Adjunctive therapy in the treatment of partial-onset seizures in adults with epilepsy

Applicant: Sepracor Inc.

Date(s): March 29, 2009

Review Priority: Standard

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Xiang Ling, Ph.D.

Concurring Reviewers: Kun Jin, Ph.D., Team Leader
Kooros Mahjoob, Ph. D.

Medical Division: Division of Neuropharmacological Drug Products, HFD-120

Clinical Team: Teresa Podruchny M.D.
Norman Hershkowitz, M.D., Ph.D., Team Leader
Russell Katz, M.D., Division Director

Project Manager: Dorothy Demczar

Keywords: log transformation, analysis of covariance, drop-outs, dose-response, multiple comparisons

Table of Contents

LIST OF TABLES.....	3
LIST OF FIGURES.....	3
1. EXECUTIVE SUMMARY	4
1.1 CONCLUSIONS AND RECOMMENDATIONS	4
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	4
1.3 STATISTICAL ISSUES AND FINDINGS	5
2. INTRODUCTION	7
2.1 OVERVIEW.....	7
2.2 DATA SOURCES	8
3. STATISTICAL EVALUATION	9
3.1. EVALUATION OF EFFICACY	9
3.1.1. <i>Study 301</i>	9
3.1.1.1. Study Design and Endpoints.....	9
3.1.1.2. Patient Disposition, Demographic and Baseline Characteristics	10
3.1.1.3. Sponsor’s Efficacy Results	12
3.1.1.4. Reviewer’s Results	15
3.1.2. <i>Study 302</i>	22
3.1.2.1. Study Design and Endpoints.....	22
3.1.2.2. Patient Disposition, Demographic and Baseline Characteristics	23
3.1.2.3. Sponsor’s Efficacy Results	26
3.1.2.4. Reviewer’s Results	29
3.1.3. <i>Supportive Study 303</i>	32
3.1.3.1. Study Design and Endpoints.....	32
3.1.3.2. Patient Disposition, Demographic and Baseline Characteristics	32
3.1.3.3. Sponsor’s Efficacy Results	34
3.1.3.4. Reviewer’s Results	36
3.2. EVALUATION OF SAFETY	38
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	39
4.1. GENDER, RACE AND AGE	39
4.2. OTHER SPECIAL/SUBGROUP POPULATIONS	40
5. SUMMARY AND CONCLUSIONS	42
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	42
5.2 CONCLUSIONS AND RECOMMENDATIONS	44
APPENDICES.....	45
APPENDIX I. STATISTICAL QUESTIONS TO SPONSOR ON SEPTEMBER 8 2009	45
APPENDIX II. STATISTICAL QUESTIONS TO SPONSOR ON OCTOBER 8 2009	46
APPENDIX III. STATISTICAL QUESTIONS TO SPONSOR ON NOVEMBER 4 2009	47
SIGNATURES/DISTRIBUTION LIST (OPTIONAL).....	ERROR! BOOKMARK NOT DEFINED.

LIST OF TABLES

Table 1. Schedule of Efficacy Assessments Conducted in the Phase III Studies	8
Table 2. Study 301: Subject Disposition	10
Table 3. Study 301: Demographic and anthropometric characteristics	11
Table 4. Study 301: Baseline Disease Characteristics (ITT Analysis Set)	12
Table 5. Study 301: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result).....	13
Table 6. Study 301: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)	13
Table 7. Study 301: Responder Analysis (Sponsor updated result).....	14
Table 8. Study 301: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed).....	20
Table 9. Study 301: FDA Analysis Results for the Primary Endpoint	21
Table 10. Study 301: FDA responder analysis	22
Table 11. Study 301: Percent Change from Baseline in seizure frequency	22
Table 12. Study 302: Subject Disposition	24
Table 13. Study 302: Demographic and anthropometric characteristics	25
Table 14. Study 302: Baseline Disease Characteristics (ITT Analysis Set)	26
Table 15. Study 302: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result).....	27
Table 16. Study 302: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)	27
Table 17. Study 302: Responder Analysis (Sponsor updated result).....	28
Table 18. Study 302: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed).....	29
Table 19. Study 302: FDA Analysis Results for the Primary Endpoint	30
Table 20. Study 302: FDA Responder Analyses	31
Table 21. Study 302: Percent Change from Baseline in seizure frequency	31
Table 22. Study 303: Subject Disposition	33
Table 23. Study 303: Demographic and anthropometric characteristics (ITT Analysis Set).....	33
Table 24. Study 303: Baseline Disease Characteristics (ITT Analysis Set)	34
Table 25. Study 303: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result).....	35
Table 26. Study 303: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)	35
Table 27. Study 303: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed).....	36
Table 28. Study 303: FDA Analysis Results for the Primary Endpoint	37
Table 29. Study 303: FDA Responder Analysis	38
Table 30. Study 303: Percent Change from Baseline in seizure frequency	38
Table 31. Treatment Effect by Gender in Pooled Population from Study 301 and Study 302	39
Table 32. Treatment Effect by Region in Study 301	40
Table 33. Treatment Effect by Region in Study 302	41
Table 34 Primary endpoint: maintenance seizure frequency	42
Table 35 Secondary endpoint: percent of responder during maintenance	42
Table 36 Percent reduction from baseline	42

LIST OF FIGURES

Figure 1. Study Treatment Procedures	8
Figure 2. Study 301: Time to withdrawal because of lack of efficacy or AEs, Kaplan-Meier survival curves (ITT population).....	11
Figure 3. Diary Card Example	16
Figure 4. Distribution of Logarithm Transformed Standardised Seizure Frequency	19
Figure 5. Study 302: Kaplan-Meier Curve of Treatment Retention Time (ITT Population).....	24

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The data seem to support the efficacy of ESL as adjunctive therapy to subjects with refractory simple or complex partial seizures. In both pivotal studies, the 800 mg/day dose resulted in statistically significantly lower standardized seizure frequency over a 12-week maintenance period compared to placebo:

$$\text{Standardized seizure frequency} = (\# \text{ of seizure} / \# \text{ of days}) * 28.$$

In the supportive phase III study, the estimate of the treatment effect for the 800 mg/day group was similar to that in the pivotal studies, although the p-value was slightly larger than 0.05 (0.0885), possibly due to the smaller ITT analysis set and larger variance.

The 1200 mg/day dose was significantly better than placebo in one of the two pivotal studies. In the other study, the results were sensitive to the handling of the early dropouts. In the analysis of the ITT population with non-conservative imputation for the early drop-outs, the p-value was 0.1143. In the supportive phase III study, the 1200 mg/day dose was marginally significant.

There is no compelling evidence that the 1200 mg/day dose provided added improvement over the 800 mg/day dose. In study 301, the 1200 mg dose group showed incremental efficacy over the 800 mg group in standardized seizure frequencies and the proportion of responders. However, the additional efficacy in the 1200 mg dose group was *not* demonstrated in study 302, where the 800 mg group had lower standardized seizure frequencies during the maintenance period.

The 400 mg/day dose was not significant in both pivotal studies.

1.2 Brief Overview of Clinical Studies

The phase III studies 2093-301, 2093-302 and 2093-303 were designed to be adequate and well-controlled trials to evaluate the efficacy and safety of BIA 2-093 as adjunctive therapy for in adults with refractory simple or complex partial seizures, with or without secondary generalization. These studies were conducted in Eastern Europe, Latin America and Western Europe/ROW. These 3 trials were similar in design: all were double-blind, randomized, placebo-controlled, parallel-group, multicentre clinical studies. However, the results from Phase III study 2093-303 were considered supportive, but not determinative, due to Good Clinical Practice concerns regarding the conduct of the study.

The design of these studies represents a standard placebo-controlled, parallel group, adjunctive therapy trial in this indication. Subjects were enrolled and entered into an 8-week Baseline Phase. Only subjects who reported ≥ 4 partial seizures in each 4-week period during the 8-week baseline period, with seizure-free period no longer than 21 days during that time frame, were to be randomized. Subjects were equally randomized to the treatment groups of placebo, 1200 mg, 800 mg, and 400 mg (400 mg group only in studies 301 and 302). After randomization, the

subjects began double-blind treatment which comprised of a 2-week titration period, and a 12-week maintenance period, with a 4-week tapering-off period in two of the three studies (301 and 303).

1.3 Statistical Issues and Findings

Study Design Issue

In this NDA, the trial participants were instructed to update seizure diary only when they experienced a seizure. As a result, “0” seizure was not recorded. Therefore, true zero seizures could not be differentiated from missing seizure data (patient did not record a seizure, missed a visit, or did not return diary card, etc.), and “no seizure data” was assumed as no occurrence of seizure events in the analysis. The worst-case-scenario analysis requested by FDA was not conducted due to the limitation of the data. See page 17 for details.

Study Conduct Issues

The sponsor has performed extensive hardcoding in the program that essentially changed the values of the variables in the database. In Study 301, 559 hard-codes (to change data in the program) were used when the sponsor generated the analysis datasets. Some hard-codes were generated from unblinded review of seizure data after data lock. In Study 302, a manually generated file was used to modify data. Upon reviewing the hardcodes, it seemed that the hardcodes were generally attempts to correct errors in the database. A sensitivity analysis with the removal of all the hardcodes associated with seizure data indicated that the result was insensitive to the hardcodes.

Hardcoding overrides the database controls in the clinical data management systems and may compromise study data integrity. Although the sponsor did not seem to intentionally mislead FDA by manipulating data, the extensive hardcoding however, indicated that the study was not well conducted and the data quality/reliability was questionable.

Statistical Analysis Issues

The sponsor specified the ITT population for the efficacy analysis as all randomised patients with at least one administration of study medication and at least one post-baseline (titration and/or maintenance period) seizure frequency assessment. However, the SAP also specified that missing values were not imputed. The majority of efficacy evaluations were performed for the maintenance period; therefore, subjects who discontinued the study during the titration period were not included in the primary and key secondary analyses. As more subjects in high dose group withdrew in early treatment phase, this analysis may be biased. The results were robust to the handling to the early drop-outs for the 800 mg/day group, but a little sensitive for the 1200 mg/day dose group as evidenced by the loss of significance in one study when a patient’s missing data after dropout during the titration period was imputed.

Natural logarithm transformation was applied to standardised seizure frequency (denoted as S) $\ln(S+4)$. The purpose of logarithm transformation was to approximate the normal distribution,

and the positive number added before the logarithm transformation was to avoid log of zero. No justification was given in the SAP for using this transformation instead of 'Ln(S+I)'. However, the Ln(S+1) approximates the normal distribution better than Ln(S+4), as confirmed by Goodness-of-Fit Tests for Normal Distribution. In addition, the statistical inference was based on LSMeans difference of the log standardize seizure frequency, which translates to the percent reduction over placebo by:

$$100*(1- \exp(\text{LSMean difference of the log standardize seizure frequency})).$$

This is approximately:

$$(\text{Treatment group frequency} - \text{Placebo group frequency})/(\text{Placebo group frequency}+4).$$

When the constant used in the logarithm transformation is large, the percent reduction over placebo is underestimated. To better approximate the normal distribution and to estimate the treatment effect, the reviewer presented results obtained using transformation 'Ln(S+I)'. The significance of the comparisons with placebo were unaffected by the change to the data transformation.

SAP stated that seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. In the updated analyses by sponsor, an additional term (the number of AEDs used at baseline) was included in the model. No justification of including this covariate was given, and this covariate did not seem to have a significant effect on the seizure frequency. The reviewer presented the results from ANCOVA models that did not include the addition term of the number of AEDs. The analysis results had minimal difference with or without this term.

2. INTRODUCTION

2.1 Overview

Class and Indication

Eslicarbazepine acetate (SEP-0002093 or BIA 2-093) is a voltage-gated sodium channel (VGSC) blocker with anticonvulsant activity. SEP-0002093 is a third-generation, single-enantiomer member of the family of first-line dibenz[b,f]azepine antiepileptic drugs (AEDs). SEP-0002093 has been developed for the adjunctive treatment of partial-onset seizures in adults with epilepsy.

History of Drug Development

Bial-Portela & Ca, S.A. (Bial) originally submitted IND 67,466 in 2006. Bial completed a transfer of ownership of the IND to Sepracor effective April 10, 2008 (Serial No. 018) and Sepracor initiated NDA preparation activities.

Eslicarbazepine acetate is the subject of a Marketing Authorization Application (MAA) submitted to the European Medicines Agency (EMA) for review via the Centralized Procedure. On February 19, 2009, the Committee for Medicinal Products for Human Use (CHMP) adopted a positive opinion, recommending the granting of a marketing authorization for eslicarbazepine acetate indicated as adjunctive therapy in adults with partial-onset seizures with or without secondary generalization.

Specific Studies Reviewed

The primary efficacy data in support of the indication are from the two pivotal Phase III studies 2093-301 and 2093-302 that includes 402 and 395 subjects respectively. The results from Phase III study 2093-303 that treated 253 subjects were submitted as a supportive, but not determinative study. Additional supportive data are available from a Phase II study that treated 143 subjects.

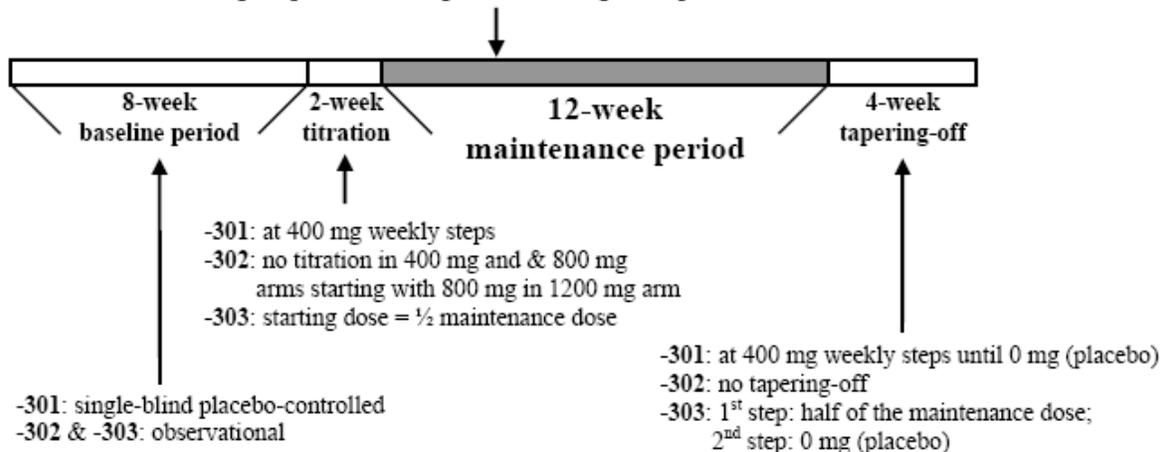
The studies were designed as multicenter, randomized, double-blind, placebo-controlled studies in adults with refractory simple or complex partial seizures, with or without secondary generalization. These studies were conducted in Eastern Europe, Latin America and Western Europe/ROW.

All Phase III studies included 2 parts. In Part 1 of each study, subjects were enrolled and entered into an 8-week Baseline Phase. Only subjects who reported ≥ 4 partial seizures in each 4-week period during the 8-week baseline period, with seizure-free period no longer than 21 days during that time frame, were to be randomized. Subjects were equally randomized to the treatment groups of placebo, 1200 mg, 800 mg, and 400 mg (400 mg group only in studies 301 and 302). After randomization, the subjects began double-blind treatment which comprised of a 2-week titration period, and a 12-week maintenance period, with a 4-week tapering-off period in two of the three studies (301 and 303). Part 2 was an open-label extension phase during which all

subjects received active treatment. The following figure summarizes the Part I of the phase III studies.

Figure 1. Study Treatment Procedures

2093-301: 4 groups – 1200 mg QD, 800 mg QD, 400 mg QD, placebo
 2093-302: 4 groups – 1200 mg QD, 800 mg QD, 400 mg QD, placebo
 2093-303: 3 groups – 1200 mg QD, 800 mg QD, placebo



Source: Sponsor’s Study Report

Table 1. Schedule of Efficacy Assessments Conducted in the Phase III Studies

	Study Period, Visit, Week					
	Baseline	Titration ^a	Maintenance ^a			Taper ^{a,b}
	V1	V2	V3	V4	V5	V6
Efficacy Parameter	-8W	0W	2W	8W	14W	18W
Review of Seizure Diary	X	X	X	X	X	X
QOLIE-31	X ^c	X			X	
MADRS	X ^c	X			X	
CGI		X			X	X

Abbreviations: V=visit, W=week, QOLIE-31=Quality of Life in Epilepsy Inventory – 31; MADRS=Montgomery-Asberg Depression Rating Scale; CGI=Clinical Global Impression scale; V1=Screening; V2=Randomization;

^a All visits during this period were to be conducted within ±3 days of the scheduled time.

^b Conducted in Studies 2093-301 and 2093-303

^c Administered at this visit for training purposes

Source: Sponsor’s ISE Report Table 1.4.3.1-1.

2.2 Data Sources

The data files are located in the following directory:

<\\Cdsub1\evsprod\NDA022416\0000\m5\datasets>

The study reports are located in the following directory:

<\\Cdsub1\evsprod\NDA022416\0000\m5\53-clin-stud-rep\535-rep-ffic-safety-stud\epilepsy\5351-stud-rep-contr>

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

3.1.1. Study 301

Study 301, titled “Efficacy and safety of BIA 2-093 as adjunctive therapy for refractory partial seizures in a double-blind, randomized, placebo-controlled, parallel-group, multicentre clinical study”, was initiated on 15 Jul 2004 (first visit of the first patient) and completed on 19 Dec 2006 (last Part I visit of the last patient). The Date of the Protocol is 8 March 2004. The final SAP version was approved and signed on 01 Feb 2006. Database lock occurred on 02 Feb 2006 and a blinded review of seizure data was conducted by Bial and the CRO right before the database lock. In response to the reviewer’s October 8, 2009 request (see appendix 2, question 1), the sponsor reported that a second review was performed on 09 Mar 2006, which was after unblinding of the data. Some hard-codes were generated from this unblinded review.

3.1.1.1. Study Design and Endpoints

This was a multicenter, randomized, double-blind, placebo-controlled study in adults with refractory simple or complex partial seizures, with or without secondary generalization. During a single-blind 8-week baseline period all patients received placebo; at the end of this period patients were randomly assigned to one of the four treatment groups: ESL 1200 mg once daily, ESL 800 mg once daily, ESL 400 mg once daily, and placebo once daily. The baseline period was followed by a double-blind 2-week titration period, a 12-week maintenance period and a 4-week tapering-off period. Part 2 of this study is a 1-year open-label extension phase.

Three regions were defined:

- Western region: Austria, Germany, Switzerland
- Central region: Croatia, Czech Republic, Hungary, Poland, Lithuania
- Eastern region: Romania, Russia, Ukraine.

Patients (or caregivers) record in a diary all seizures by date, time of occurrence, seizure type, and number throughout the study. Patients were instructed to update diaries every time they experience a seizure. Diaries were distributed and collected at each study visit.

The primary efficacy endpoint: standardized seizure frequency per 4 weeks over the 12-week maintenance period.

Secondary efficacy endpoints:

- Proportion of patients with a 50% or greater reduction in seizure frequency during the 12-week maintenance period compared with the 8-week baseline period (responders);
- Seizure frequency per week for each week of the baseline, titration, maintenance and tapering-off periods;
- Distribution of seizure reduction (< 50%, 50–75%, or >75% seizure reduction);
- Proportion of seizure-free patients (100% seizure reduction);

- Proportion of patients with a 25% exacerbation in seizure frequency compared to baseline;
- Seizure frequency by seizure type;
- Seizure frequency as a function of BIA 2-194 plasma levels at Visit 5;
- Treatment retention time (time to withdrawal due to lack of efficacy or adverse events) during Part I of the study;
- Proportion of patients remaining on treatment for the duration of Part I of the study;
- Clinical global impressions (CGIs);
- Responses to the Quality of Life in Epilepsy Inventory-31 (QOLIE-31);
- Symptoms of depression (based on the Montgomery Asberg Depression Rating Scale [MADRS]).

3.1.1.2. Patient Disposition, Demographic and Baseline Characteristics

A total of 468 subjects were enrolled and 402 subjects were randomized into this study. A total of 330 patients completed Part I of the study. The proportion of randomised patients who discontinued the study prematurely was highest (30.4%, corresponding to 31 patients) in the ESL 1200 mg group. The most frequent reason for discontinuation in this group was “occurrence of unacceptable AE” being applicable to 18 patients (17.6%). In the ESL 1200 mg group, withdrawals (due to lack of efficacy or AEs) predominantly occurred during the early treatment phase.

Table 2. Study 301: Subject Disposition

Disposition	Placebo	Eslicarbazepine Acetate Dose Group			Total
		400 mg	800 mg	1200 mg	
Randomized	102	100	98	102	402
Completed the Study	84 (82.4)	90 (90.0)	85 (86.7)	71 (69.6)	330 (82.1)
Withdrew Prematurely	18 (17.6) ^a	10 (10.0)	13 (13.3)	31 (30.4) ^a	72 (17.9) ^a
Withdrew during/end of Baseline	1 (1.0)	2 (2.0)	2 (2.0)	7 (6.9)	12 (3.0)
Withdrew during/end of Titration	5 (4.9)	2 (2.0)	4 (4.1)	13 (12.7)	24 (5.9)
Withdrew during/end of Maintenance	10 (9.8)	6 (6.0)	7 (7.1)	10 (9.8)	33 (8.2)
Withdrew during/end of Taper	0	0	0	0	0
Reasons for Withdrawal ^b					
Unacceptable Adverse Event	3 (2.9)	4 (4.0)	8 (8.2)	18 (17.6)	33 (8.2)
Withdrawal of Consent	10 (9.8)	4 (4.0)	4 (4.1)	10 (9.8)	28 (7.0)
Subject Non-compliance	3 (2.9)	1 (1.0)	1 (1.0)	1 (1.0)	6 (1.5)
Lack of Efficacy	0	0	0	2 (2.0)	2 (0.5)
Investigator Discretion	0	0	1 (1.0)	1 (1.0)	2 (0.5)
Protocol violation	1 (1.0)	0	0	1 (1.0)	2 (0.5)
Pregnancy	0	0	1 (1.0)	0	1 (0.2)
Other	4 (3.9)	2 (2.0)	1 (1.0)	3 (2.9)	10 (2.5)

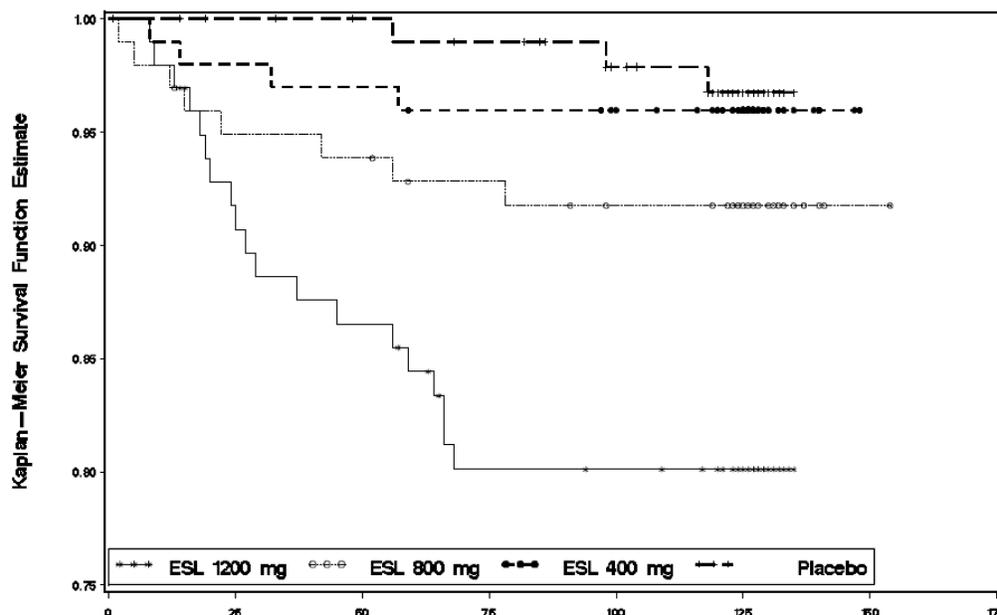
a Includes 2 subjects from the placebo group and 1 subject from the ESL 1200 mg group who finished the tapering-off period but who were reported not to have completed Visits 3 through 6 as scheduled; a reason for discontinuation was therefore specified.

b A subject may have more than 1 reason for discontinuation.

Reference: CSR 2093-301 Part 1, in-text [Table 2](#) and Section 15, [Table 16](#).

Source: Sponsor’s ISE Report Table 2.1.1-1.

Figure 2. Study 301: Time to withdrawal because of lack of efficacy or AEs, Kaplan-Meier survival curves (ITT population)



Best Available Copy

Source: Sponsor’s Study Report Figure 25.

In original CSR, five randomized patients were not eligible for the ITT population because of no post-baseline seizure assessment (4 in ESL 1200 mg group and 1 in ESL 400 mg group). *However, based on sponsor updated analysis data set, 7 patients did not have post-baseline seizure assessment.*

With respect to the demographic characteristics of the ITT population there were no relevant differences between the treatment groups. On average, the patients’ age was lowest in the placebo group (mean±SD: 37.0±11.93 years) and highest in the ESL 800 mg group (41.3±12.04 years). The gender distribution varied between 55% males in the ESL 800 mg group and 44% males in the ESL 1200 mg group. All study patients were Caucasians.

Table 3. Study 301: Demographic and anthropometric characteristics

Parameter	Statistic/Class	Placebo (N=102)	Eslicarbazepine Acetate Dose Group			Total (N=397)
			400 mg (N=99)	800 mg (N=98)	1200 mg (N=98)	
Age (years)	N	102	99	98	98	397
	Mean (SD)	37.0 (11.93)	37.9 (11.47)	41.3 (12.04)	38.3 (11.75)	38.6 (11.86)
	Min, Max	18.3, 68.6	18.0, 71.0	18.7, 75.6	19.0, 65.2	18.0, 75.6
Sex	% Male	47	51	55	44	49
	% Caucasian	100	100	100	100	100
BMI (kg/m ²)	N	102	99	98	98	397
	Mean (SD)	24.5 (4.48)	24.4 (4.45)	24.5 (4.63)	24.4 (4.16)	24.5 (4.42)
	Min, Max	16.0, 39.9	15.2, 40.4	16.0, 43.4	16.6, 38.2	15.2, 43.4

Abbreviations: SD=standard deviation; BMI=body mass index; Min=minimum; Max=maximum

Source: Sponsor’s ISE Report Table 2.1.2-1.

In the ITT population there were no relevant group-specific differences regarding the seizure frequency and the number of concomitant AEDs per patient at baseline.

Table 4. Study 301: Baseline Disease Characteristics (ITT Analysis Set)

Parameter	Statistic/Class	Placebo (N=102)	Eslicarbazepine Acetate Dose Group		
			400 mg (N=99)	800 mg (N=98)	1200 mg (N=98)
Duration of Epilepsy (years)	N	102	99	98	98
	Mean (SD)	19.4 (12.57)	21.0 (11.70)	23.1 (13.50)	20.4 (11.85)
	Min, Max	1.0, 54.0	1.0, 55.2	1.1, 53.0	1.1, 45.8
Most Common ^a Etiology of Epilepsy	%Other	39	37	39	35
	%Cranial trauma	20	26	19	15
	%Idiopathic	25	15	15	16
	%Infection	6	15	13	10
Seizure Frequency at Visit 2 ^b	N	102	99	98	98
	Mean (SD)	12.4 (17.94)	11.4 (9.74)	11.2 (11.21)	11.6 (15.92)
	Median	6.7	7.5	7.0	7.5
	Min, Max	2.0, 153.5	2.5, 55.5	3.0, 70.5	3.6, 141.5
Concomitant AED at Baseline	%Taking Any AED	100	100	100	100
	%Taking 1 AED	33	39	32	39
	%Taking 2 AEDs	66	60	68	61
Most Common ^a Types of AED	%Carbamazepine	62	56	60	56
	%Lamotrigine	26	24	27	28
	%Valproic acid	28	26	22	26
	%Topiramate	16	9	19	11

Abbreviations: SD=standard deviation; AED=antiepileptic drug; Min=minimum; Max=maximum

a Reported in $\geq 10\%$ of the total ITT analysis set

b Standardized to 4 weeks.

Source: Sponsor's ISE Report Table 2.1.2-2.

3.1.1.3. Sponsor's Efficacy Results

Sponsor's primary efficacy results

The SAP specified efficacy ITT data set as all randomised patients with at least one administration of study medication and at least one post-baseline seizure frequency assessment. The SAP also stated that no imputation methods would be used for subjects without data in the maintenance period. As a result, the primary analysis was performed on patients with at least one seizure frequency assessment during maintenance period. This deviated from the ITT principle.

Seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. Dunnett's multiple comparison procedure was used for the comparison of the active treatment means to the placebo mean. Natural logarithm transformation was applied to standardised seizure $\text{Ln}(\text{standardized seizure frequency}+4)$. The standardized seizure frequency for a period is calculated by $(\text{number of seizures/days in the period} \times 28)$. The original result was presented in the table below:

Table 5. Study 301: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result)

Standardized Seizure Frequency per 4 weeks	Placebo (N=102)	Eslicarbazepine Acetate Dose Group		
		400 mg (N=99)	800 mg (N=98)	1200 mg (N=98)
N	99	97	94	94
LS Mean (back-transformed)	7.64	6.73	5.66	5.35
[95% CI on the Mean]	[6.78, 8.58]	[5.93, 7.60]	[4.92, 6.45]	[4.63, 6.12]
Log LS Mean Difference to Placebo		-0.0812	-0.1869	-0.2196
ANCOVA p-value ^a		0.3332	0.0028	0.0003

a p-value from Dunnett's multiple comparison procedure for the comparison of the active treatment means to the placebo mean.

Source: Sponsor's ISE Report Table 2.1.3-1.

The ANCOVAs of reduction in seizure frequency per 4 weeks for the PP population during the 12-week maintenance period and for the ITT and PP populations during the 2-week titration and 12-week maintenance periods showed results similar to those obtained with the primary efficacy analysis.

The primary analysis was updated upon FDA Biostat request to correct several errors in deriving the efficacy variable. The updated result was similar.

Table 6. Study 301: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
N	99	97	93	92
LSmean (SE)	7.5 (0.67)	6.7 (0.60)	5.6 (0.58)	5.4 (0.56)
95% CI	6.3, 8.9	5.6, 8.0	4.6, 6.9	4.4, 6.6
Log Difference in LS Mean (SE)		-0.07 (0.055)	-0.18 (0.055)	-0.20 (0.056)
95% CI for Difference in LS Mean		-0.20, 0.06	-0.31, -0.05	-0.33, -0.07
p-value		0.4067	0.0041	0.0009

Source: Sponsor's response to October 8, 2009 Request for information and is confirmed by FDA reviewer. Without imputation, baseline AED as covariate

Secondary efficacy endpoint results

In the SAP for this study, the proportion of responders (defined as subjects with a $\geq 50\%$ reduction in seizure frequency from baseline) over the 12-week maintenance period was analysed by using a Cochran-Mantel-Haenszel (CMH) test. Subjects who did not enter the particular period were considered as non-responders per SAP.

The submitted analysis using the updated datasets, however, was different from the original specification. The comparison was done using continuity-adjusted Chi-Square test. Subjects who did not have seizure data during the maintenance were not included. According to this analysis, the responder rate was significantly higher in the ESL 1200 mg group (44.6%) and the ESL 800 mg group (35.5%) than the placebo group (20.2%) for the ITT population (unadjusted $p = 0.0006$ and 0.0274 , respectively). Comparison of the ESL 400 mg group to placebo was not significant ($p=0.6738$). The tests were not adjusted for multiplicity.

Table 7. Study 301: Responder Analysis (Sponsor updated result)

Responder	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
n/N (%) ^a	20/99 (20.2)	23/97 (23.7)	33/93 (35.5)	41/92 (44.6)
Chi-square p-value ^b		0.6738	0.0274	0.0006
Odds Ratio		1.23	2.17	3.18
95% CI		0.62, 2.42	1.14, 4.16	1.67, 6.02

a. n/N=number of responders/number of subjects with seizure data in the maintenance period.

b. p-value from pairwise test of each active treatment group compared to placebo.

Source: Sponsor's response to October 8, 2009 Request for information.

Sponsor's covariate assessment and subgroup analyses

The following subgroup analyses and covariate assessment were from the Sponsor's CSR and were not based on the updated datasets. Analyses of seizure frequency were performed using the ANCOVA model for the primary efficacy assessment but with region and the treatment-by-region interaction added to the model. In the ITT population, the analysis showed that region had no effect on seizure frequency and that there was no interaction between region and treatment. In the PP population, the treatment-by-region interaction was significant during the 12-week maintenance period ($p = 0.0450$), but no effect of region was found ($p = 0.2676$).

The analysis of seizure frequency with carbamazepine use and treatment-by-carbamazepine use as factors showed that treatment-by-carbamazepine interaction was not significant but the use of carbamazepine had a significant effect on seizure frequency during the 12-week maintenance period ($p = 0.0318$), but not during the 2-week titration plus 12-week maintenance period for the ITT population and not for the PP population.

As regards the use of concomitant AEDs at or after Visit 2 there were some remarkable differences between the western, the central and the eastern region. In the central region a higher proportion of patients took two AEDs (80.6% and 65.5% respectively) than in the eastern region (43.1%). The most frequently prescribed AEDs were carbamazepine (74.3%) in the eastern region; carbamazepine (49.8%), lamotrigine (36.8%) and valproic acid (33.3%) in the central region; and levetiracetam (48.3%), lamotrigine (48.3%) and carbamazepine (27.6%) in the western region.

Additional analyses with the following covariates for the primary endpoint were performed: age, gender, use of concomitant non-AEDs, number of concomitant AEDs, and epilepsy duration. No significant effects were observed for those covariates.

Sponsor's Efficacy Summary

ESL 800 and 1200mg/day treatment groups were statistically superior to the placebo group in seizure frequency reduction at maintenance endpoint (800mg/day p -value = 0.0041; 1200mg/day p -value=0.0009). The percent reduction in seizure frequency over placebo was 16.5% and 18.1% for ESL 800mg/day and 1200mg/day, respectively. Similar positive findings were observed in the statistical analysis of responders, which is defined as subjects with at least 50% of seizure reduction at maintenance endpoint. The $\geq 50\%$ responder rates for placebo, ESL

800mg/day and 1200mg/day were 22%, 35% and 45%, respectively. The p-values for ESL 800mg/day and 1200mg/day when compared with placebo from CMH test were 0.0274 and 0.0006, which indicate the two dose groups are more likely than placebo group to have 50% responders. Results were robust with consistent conclusions drawn from the analyses conducted on both the ITT and the PP analysis sets.

The percent reduction in partial seizure frequency over placebo was 6.8% (p-value = 0.4067) for ESL 400mg/day indicating a numerically greater but not statistically significant difference. The 50% responder rate was 24% (p-value = 0.6738) for ESL 400mg/day, which was similar to the placebo group.

Data from this trial in the sponsor's opinion demonstrate that ESL 800mg/day and 1200mg/day is an effective treatment.

In addition, as evidenced by analysis of data from the tapering-off period in Study 2093-301, continuation of treatment is required to maintain effect. In the 1-year treatment extension period of the pivotal Study 2093-301, the efficacy of SEP-0002093 was maintained and no tolerance to the observed effects was evident (details in sponsor module 2.5 section 4.5).

3.1.1.4. Reviewer's Results

The reviewer has identified the following issues in sponsor's analyses.

Extensive hard-codes

Based on sponsor's response to the reviewer's October 8, 2009 request (appendix 2), 559 hard-codes were used when the sponsor generated the analysis datasets. Hardcodes essentially changed the values of the variables in the database. Upon reviewing the hardcodes, it seemed that the hardcodes were generally attempts to correct errors in the database. For example, the same seizure was reported on two different diaries and the programming manually removes one of them.

The percent of seizure records that were changed is 3.3%, and the subjects with at least one change in their seizure counts at either baseline or during the maintenance period were evenly distributed across the treatment groups. A sensitivity analysis with the removal of all the hardcodes associated with seizure data indicated that the result was insensitive to the hardcodes.

The extensive hardcoding, however, indicated that the study was not well conducted and the data quality/reliability was questionable.

Unblinded seizure data review

Based on sponsor's response to Biostat October 8, 2009 request, a blinded review of seizure data was conducted by Bial and the CRO on 01 Feb 2006, and a second review was performed on 09 Mar 2006, which was after unblinding of the data. Some hard-codes were generated from this unblinded review. *In the original submission, the sponsor only mentioned the blinded review, not*

the unblinded review. As the unblinded review may potentially bias the study result, the sponsor conducted a sensitivity analysis removing the hardcodes associated with seizure data. No difference in data interpretation or conclusions was observed.

Missing seizure data

The seizure frequency was calculated as (Number of seizures *28/ number of days). The number of days should only include days for which seizure information was provided, and days when patients fail to record seizures should be excluded. In seizure trials, the diary cards are usually required to be updated on a daily basis and calculating the seizure frequency is straight forward. In this NDA, however, the trial participants were instructed to update seizure diary only when they experienced a seizure. As a result, “0” seizure was not recorded. Therefore, true zero seizures could not be differentiated from missing seizure data (patient did not record a seizure, missed a visit, or did not return diary card, etc.). The following assumptions/methods were applied in this submission:

- a. “No seizure data” was assumed as no occurrence of seizure events. Unreturned diary cards, and unfilled diary cards, therefore, were treated as no seizures, although it was possible that the patient failed to record the seizures that he/she experienced.
- b. The last seizure diary return date was used to as the end date for calculating the duration of the period. This assumed that a patient would have recorded any seizures that occurred before/on the date that the subject returned the last diary card during a study visit.

As an example, the following were the post-baseline diary cards for subject 112-90327. The dataset “Diaryd” indicated that the subject had baseline frequency of 14.71 per 4 weeks and started maintenance phase on 06DEC2004. The diary cards showed 2 seizures during maintenance period and 18 seizures (occurred within one month) during tapering-off period. Assuming empty diary cards indicating no seizures, the resulting seizure frequency for the maintenance period was 0.67.

Figure 3. Diary Card Example



A worst-case-scenario analysis was performed by sponsor to assess the effect of the part of missing data that were caused by unreturned diary cards. The FDA requested worst-case-scenario analysis (see appendix Statistical Questions to Sponsor on November 4 2009 Statistical, question 3) was not doable because the period of time for which the diaries were missing was undeterminable.

In this analysis, for each diary card a subject failed to return, the number of expected seizures was determined using a worst-case scenario according to the subject's treatment group as follows:

Period	Placebo Subjects	ESL Subjects
Baseline	total number of seizures reported during the maintenance period divided by the number of diary cards returned during the maintenance period	0
Maintenance	0	total number of seizures reported during the baseline period divided by the number of diary cards returned during the baseline period

For each subject with missing diary cards, a new estimate of the total number of seizures within the treatment period was calculated as:

(# seizures reported on returned diary cards) + (# missing diary cards x the worst-case estimate of the number of seizures per diary card).

The analysis with worst-case imputation of missing data was still favorable. P-value changed from 0.0041 to 0.0599 for the 800 mg group and from 0.0009 to 0.0144 for the 1200 mg group.

Study population

The sponsor specified the ITT population for the efficacy analysis as all randomised patients with at least one administration of study medication and at least one post-baseline (*titration and /or maintenance period*) seizure frequency assessment. However, the primary analysis did not impute seizure data, hence only included patients with seizure data during the *maintenance* period, which deviated from the ITT principle. As more subjects in high dose group withdrew in early treatment phase, this analysis may be biased.

The sponsor also specified a secondary analysis, in which the maximum seizure frequency during the baseline or the titration period was used as the seizure frequency during the maintenance period. This approach was used for the following reason. An underreporting of seizures was expected for drop-out during the titration period. As more subjects from the high dose group dropped out during titration period, the seizure frequency may be underestimated to a larger extent for the high dose group than the placebo group. The usual way of handling drop-outs (use the seizure frequency during the titration period for the maintenance period) may favor the high dose group when assessing the drug efficacy.

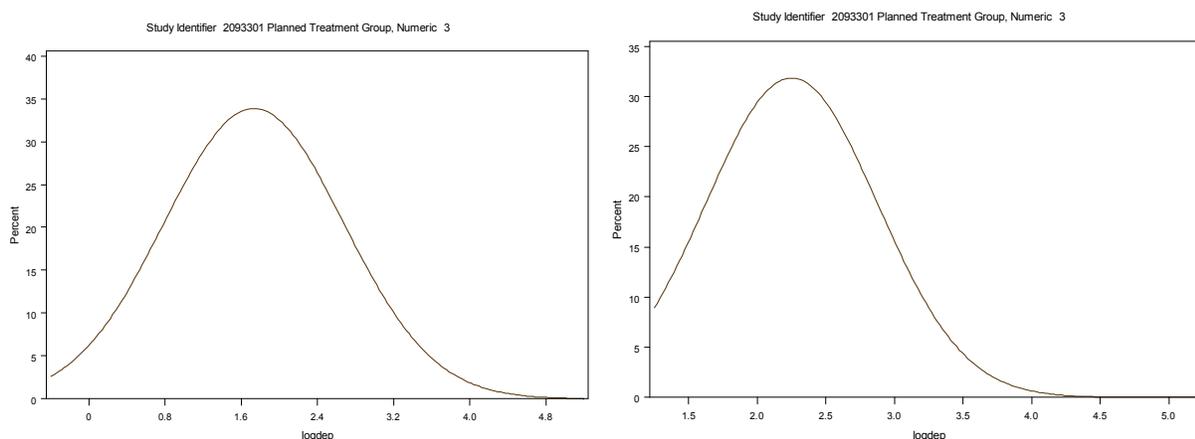
For example, subject 90506 had baseline seizure frequency of 20.5 per 4 weeks. This subject dropped out during titration period and no seizure was reported for titration period (titration period started on 03FEB2005 and last seizure diary return date is 17FEB2005), resulting in a seizure frequency of 0 for titration period. The baseline seizure frequency of 20.5 per 4 wks was imputed for the maintenance period.

In this secondary analysis, all randomized subjects (including those without any post-baseline seizure assessments) were included, which was not the ITT population either.

Data transformation

Natural logarithm transformation was applied to standardised seizure frequency (denoted as S) $\ln(S+4)$. The purpose of logarithm transformation was to approximate the normal distribution, and the positive number added before the logarithm transformation was to avoid log of zero. No justification was given in the SAP for using this transformation instead of ' $\ln(S+1)$ '. However, the $\ln(S+1)$ approximates the normal distribution better than $\ln(S+4)$, as confirmed by Goodness-of-Fit Tests for Normal Distribution. The follow plots showed the distribution of the logarithm transformed standard seizure frequency for the high dose group during maintenance period. The left one is for $\ln(S+1)$ and the right one is for $\ln(S+4)$.

Figure 4. Distribution of Logarithm Transformed Standardised Seizure Frequency



Best Available Copy

Source: FDA

The statistical inference was based on LSMean difference of the log standardize seizure frequency, which translates to the percent reduction over placebo by $100 \cdot (1 - \exp(\text{LSMean difference of the log standardize seizure frequency}))$. This is approximately $(\text{Treatment group frequency} - \text{Placebo group frequency}) / (\text{Placebo group frequency} + 4)$. When the constant used in the logarithm transformation is large, the percent reduction over placebo is underestimated. The underestimation is greater especially when many patients with low seizure frequency were enrolled. Note that the protocol required that only subjects who reported ≥ 4 partial seizures in each 4-week period during the 8-week baseline period, with seizure-free period no longer than 21 days during that time frame, were to be randomized. This was necessary to detect both decreases and increases in seizure frequency. However, there were 12 ITT subjects with baseline seizure frequency less than 4.

In addition, the Standard Error of the LSMean difference of the log standardize seizure frequency is underestimated too. This can be proved analytically using the Delta Method. Therefore, the impact of the transformation on the test result depends on the extent of the underestimation of both the point estimation and the SE.

ANCOVA model

SAP stated that seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. In the updated analyses by sponsor, an additional term (the number of AEDs used at baseline) was included in the model. No justification of including this covariate was given, and this covariate not did seem to have a significant effect on the seizure frequency.

Results

Descriptive Summary

Observed median relative reduction in seizure frequency for the ITT population during the 12-week maintenance period was 16.7%, 25.8%, 36.5% and 41.9% in the placebo group, ESL 400 mg, ESL 800 mg and ESL 1200 mg groups respectively (Table 8).

Table 8. Study 301: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed)

	Eslicarbazine Acetate Dose Group			
	Placebo	400 mg	800 mg	1200 mg
Baseline	6.7	7.5	7.0	7.4
Titration	7.9	7.5	6.0	6.6
Maintenance	6.7	5.7	4.9	4.7
Percent Change from Baseline to Maintenance (%)	-16.7	-25.8	-36.5	-41.9

Source: FDA reviewer

Analyses of Primary Endpoint

In the reviewer analyses:

- Only baseline seizure frequency and treatment group were included in the ANCOVA model
- Log standardised seizure was calculated as $\text{Ln}(\text{standardized seizure frequency} + 1)$ to have a better estimate of the percent reduction over placebo
- For ITT population of subject with post-baseline seizure information, two methods were used to impute missing seizure frequency during the maintenance period
 - Conservative Imputation: The maximum seizure frequency during the baseline or the titration period. This may be conservative.
 - Non-conservative Imputation: Seizure frequency during titration period. This method is not conservative, and is the primary analysis method in NDA 22,253 (Lacosamide).

The results showed that the 800 and 1200 mg/day doses had statistically significantly lower seizure frequencies than placebo during assessment period. The results were robust to the handling of dropouts (Table 9). Based on the analysis using non-conservative imputation on ITT population, the percent reduction over placebo calculated by $100 * (1 - \exp(\text{LSMean difference of the log standardize seizure frequency}))$ were 9.5%, 21.8% and 27.3% for 400mg, 800mg and 1200mg groups respectively. The dose response appeared to be monotone.

Table 9. Study 301: FDA Analysis Results for the Primary Endpoint

	Eslicarbazepine Acetate Dose Group			
	Placebo	400 mg	800 mg	1200 mg
Completers (with maintenance assessment)				
N	99	97	93	92
LSmean (SE)	6.8 (0.47)	6.1 (0.44)	5.1 (0.38)	4.6 (0.35)
95% CI	6.0, 7.8	5.3, 7.0	4.4, 5.9	4.0, 5.4
Log Difference in LSMean (SE)		-0.10 (0.086)	-0.25 (0.087)	-0.33 (0.087)
95% CI for Difference in LSMean		-0.30, 0.11	-0.45, -0.04	-0.54, -0.13
p-value		0.5517	0.0128	0.0004
ITT population (Conservative Imputation)				
N	102	98	98	97
LSmean (SE)	7.0 (0.48)	6.2 (0.44)	5.3 (0.38)	4.8 (0.36)
95% CI	6.1, 8.0	5.4, 7.1	4.6, 6.1	4.2, 5.6
Log Difference in LSMean (SE)		-0.11 (0.086)	-0.23 (0.086)	-0.31 (0.086)
95% CI for Difference in LSMean		-0.31, 0.10	-0.44, -0.03	-0.51, -0.11
p-value		0.4715	0.0181	0.0010
ITT population (Non-conservative Imputation)				
N	102	98	98	97
LSmean (SE)	6.9 (0.48)	6.2 (0.44)	5.2 (0.38)	4.8 (0.36)
95% CI	6.0, 7.9	5.3, 7.1	4.5, 6.0	4.1, 5.5
Log Difference in LSMean (SE)		-0.10 (0.086)	-0.25 (0.086)	-0.32 (0.086)
95% CI for Difference in LSMean		-0.30, 0.10	-0.45, -0.04	-0.52, -0.12
p-value		0.5136	0.0125	0.0007

Source: FDA

Site 112 was identified to have compliance issues by FDA inspection. Analyses excluding site 112 showed similar results.

Analyses of Secondary Endpoints

Percent of responder

The analyses of responders were conducted for ITT population of subject with post-baseline seizure information. Subjects who did not have seizure data during the maintenance period were treated in two ways:

- I. They were considered as non-responders, per sponsor SAP;
- II. Responses during titration were used.

P-values were computed from 2 tests:

- a. CMH test, per sponsor SAP;
- b. Continuity-Adjusted Chi-square test, using sponsor's ISE analysis method.

The result of this secondary endpoint supported the primary analysis (Table 10).

Table 10. Study 301: FDA responder analysis

Responder	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
I. Subjects w/o maintenance data as non-responder				
n/N (%) ^a	20/102 (19.6)	23/ 98 (23.5)	33/ 98 (33.7)	41/ 97 (42.3)
CMH p-value ^b		0.5074	0.0246	0.0005
Chi-square p-value ^b		0.6225	0.0364	0.0009
Odds Ratio		1.26	2.08	3.00
95% CI		0.64, 2.47	1.09, 3.96	1.59, 5.66
II. Impute using titration				
n/N (%) ^a	20/102 (19.6)	23/ 98 (23.5)	34/ 98 (34.7)	42/ 97 (43.3)
CMH p-value ^b		0.5074	0.0166	0.0003
Chi-square p-value ^b		0.6225	0.0249	0.0006
Odds Ratio		1.26	2.18	3.13
95% CI		0.64, 2.47	1.15, 4.14	1.66, 5.89

a. n/N=number of responders/number of subjects with seizure data in the maintenance period.

b. Unadjusted p-value from pairwise test of each active treatment group compared to placebo.

Source: FDA.

Percent change from baseline in seizure frequency

This was not one of the secondary endpoint specified in the study protocol. However, it was included in the submission. Analysis was conducted in a similar manner to the primary endpoint. Below was the result for ITT population with non-conservative imputation (using titration data for patients without maintenance data). Results indicate that both 800 mg and 1200 mg dose groups showed significantly greater reductions from baseline compared to the placebo group in standardized seizure frequency. The percent reduction was similar for the 1200 mg dose group and 800 mg dose group (Table 11).

Table 11. Study 301: Percent Change from Baseline in seizure frequency

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
N	102	98	98	97
LSmean (SE) (%)	-7.7 (5.87)	-15.9 (5.98)	-28.4 (5.98)	-29.6 (6.01)
95% CI (%)	-19.2, 3.8	-27.7, -4.2	-40.2, -16.7	-41.4, -17.7
Difference in LSMean (SE) (%)		-8.24 (8.382)	-20.74 (8.383)	-21.88 (8.402)
95% CI for difference in LSMean(%)		-28.02, 11.54	-40.52, -0.95	-41.71, -2.05
p-value		0.6391	0.0373	0.0262

Source: FDA

3.1.2. Study 302

3.1.2.1. Study Design and Endpoints

Study 302, titled “Efficacy and safety of BIA 2-093 as adjunctive therapy for refractory partial seizures in a double-blind, randomized, placebo-controlled, parallel-group, multicentre clinical study”, was initiated on 01 Sep 2004 and completed on 19 Dec 2006. The original protocol was issued on 08 Mar 2004 and was amended 3 times. Amendment 1 was issued on 17 Aug 2004 and

applied only to the Site 352. To facilitate patient recruitment, Amendment 2 was issued on 26 Oct 2005 to allow the recruitment of patients with up to 3 concomitant AEDs, rather than 2 AEDs as originally specified in the protocol. Amendment 3 was issued on 12 Apr 2006 to add an optional extension period after Part II of the present study. The Date of the SAP was 22 Dec 2006 and the final signature was collected on 02 Jan 2007. The study unblinding date is 14 Feb 2007, based on the sponsor's reply to the reviewer's inquiry.

This was a multicenter, randomized, double-blind, placebo-controlled study consisted of an 8-week baseline period, followed by a double-blind 2-week titration period and a 12-week maintenance period. At the end of the baseline period, patients were be randomly assigned to one of 4 treatment groups: BIA 2-093 1200 mg once-daily, BIA 2-093 800 mg once-daily, BIA 2-093 400 mg once-daily, and Placebo. Patients attended the study clinic for up to 5 scheduled visits during Part I. The study schedule was similar to Study 301 except that there was no tapering-off period. Part II was a 22-week open-label extension phase.

Three regions were defined: Europe (Belgium, Denmark, Germany, the Netherlands, Portugal, Romania, Spain, Sweden, and the United Kingdom), South America (Argentina and Brazil), and Australia and South Africa.

The primary and secondary efficacy endpoints were the same as study 301.

3.1.2.2. Patient Disposition, Demographic and Baseline Characteristics

A total of 503 subjects were screened at 46 sites in 13 countries and 395 subjects were randomized into this study. A total of 327 patients completed Part I of the study. The proportion of randomized patients who discontinued the study prematurely was highest (31%) in the ESL 1200 mg group. The most frequent reason for discontinuation in this group was "occurrence of unacceptable AE" being applicable to 25 (25.5%) patients. A higher proportion of subjects in the ESL 800 mg and 1200 mg groups discontinued during the titration phase (8% and 6%, respectively) compared to the placebo and ESL 400 mg groups (0%).

Table 12. Study 302: Subject Disposition

Disposition	Placebo	Eslicarbazepine Acetate Dose Group			Total
		400 mg	800 mg	1200 mg	
Randomized	100	96	101	98	395
Completed the Study	94 (94.0)	84 (87.5)	81 (80.2)	68 (69.4)	327 (82.8)
Withdrew Prematurely	6 (6.0)	12 (12.5)	20 (19.8)	30 (30.6)	68 (17.2)
Withdrew end of Baseline	0	0	1 (1.0)	0	1 (2.5)
Withdrew during Titration	0	0	8 (7.9)	6 (6.1)	14 (3.5)
Withdrew during/end of Maintenance	6 (6.0)	12 (12.5)	11 (10.9)	24 (24.5)	53 (13.4)
Reasons for Withdrawal ^a					
Unacceptable Adverse Event	3 (3.0)	10 (10.4)	16 (15.8)	25 (25.5)	54 (13.7)
Withdrawal of Consent	1 (1.0)	2 (2.1)	2 (2.0)	3 (3.1)	8 (2.0)
Subject Non-compliance	2 (2.0)	1 (1.0)	0	4 (4.1)	7 (1.8)
Lack of Efficacy	1 (1.0)	2 (2.1)	0	0	3 (0.8)
Protocol violation	0	1 (1.0)	1 (1.0)	1 (1.0)	3 (0.8)
Investigator Discretion	0	0	1 (1.0)	1 (1.0)	2 (0.5)
Other	0	1 (1.0)	2 (2.0)	1 (1.0)	4 (1.0)

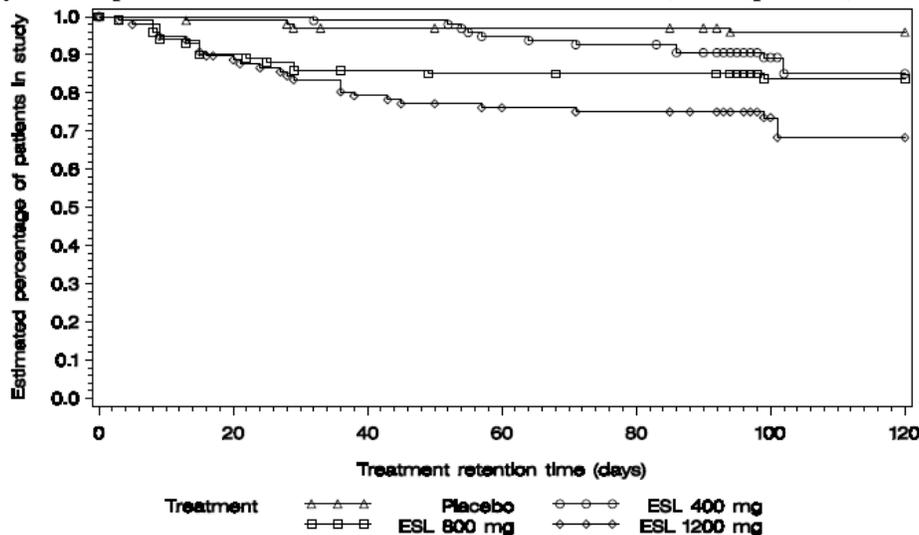
a A subject may have more than 1 reason for discontinuation.

Reference: CSR 2093-302, Section 16.2.1, Listing 16.2.1-2

Source: Sponsor’s ISE Report Table 2.2.1-1.

Treatment retention time during Part I of the study was defined as the day of the termination from the double-blind assessment period (if primary reason for withdrawal was unacceptable AE, exacerbation of seizures, or withdrawal of consent). Those patients completing the double-blind assessment period or withdrawing with a primary withdrawal reason other than unacceptable AE, exacerbation of seizures, or withdrawal of consent were censored at the day of the termination or completion visit.

Figure 5. Study 302: Kaplan-Meier Curve of Treatment Retention Time (ITT Population)



Best Available Copy

Source: Sponsor’s study Report Figure 11-3.

The ITT data set consisted of all randomized patients with at least one administration of study medication and at least one post-baseline seizure frequency assessment. Of the 395 patients were randomized, 2 treated patients had no post-baseline seizure frequency assessment and were

therefore excluded from the ITT population of 393 patients. *However, based on the updated dataset, only 388 patients had post-baseline seizure data.*

Note that in the ITT population, 35 subjects including 10%, 9%, 6%, and 10% of subjects in the placebo and ESL 400 mg, 800 mg and 1200 mg groups, respectively had baseline seizure frequency <4 per 4 weeks, a major protocol violation.

With respect to the demographic characteristics of the ITT population there were no relevant differences between the treatment groups. The gender distribution varied between 41% males in the ESL 400 mg group and 54% males in the ESL 1200 mg group. The majority of subjects enrolled were Caucasian (87.5%).

Table 13. Study 302: Demographic and anthropometric characteristics

Parameter	Statistic/Class	Placebo (N=100)	Eslicarbazepine Acetate Dose Group			Total (N=393)
			400 mg (N=96)	800 mg (N=100)	1200 mg (N=97)	
Age (years)	N	100	96	100	97	393
	Mean (SD)	36.7 (12.17)	37.6 (11.19)	36.6 (12.58)	36.9 (11.60)	36.9 (11.87)
	Min, Max	18.0, 68.0	18.0, 67.0	18.0, 65.0	18.0, 69.0	18.0, 69.0
Sex	% Male	52	41	51	54	49
Race	% Caucasian	87	91	90	82	88
BMI (kg/m ²)	N	100	96	100	97	393
	Mean (SD)	25.2 (4.68)	24.8 (5.00)	24.7 (4.36)	25.3 (4.54)	25.0 (4.64)
	Min, Max	16.7, 38.5	17.7, 41.1	17.0, 39.7	18.5, 40.2	16.7, 41.1

Source: Sponsor's ISE Report Table 2.2.2-1.

Median seizure frequency during the 8-week baseline period, standardized to 4-weeks, was 7.4, 8.2, 9.1, and 9.3 seizures per 4 weeks in the placebo and ESL 400 mg, 800 mg and 1200 mg groups, respectively. A majority of all subjects regularly took 2 concomitant AEDs. The most common concomitant AEDs were Carbamazepine, Valproic acid and Lamotrigin. Study 302 appears to have evaluated a more severely ill population with a higher baseline seizure frequency and taking more AEDs.

Table 14. Study 302: Baseline Disease Characteristics (ITT Analysis Set)

Parameter	Statistic/Class	Placebo (N=100)	Eslicarbazepine Acetate Dose Group		
			400 mg (N=96)	800 mg (N=100)	1200 mg (N=97)
Duration of Epilepsy (years)	N	100	95	100	97
	Mean (SD)	25.4 (13.06)	24.7 (11.52)	22.6 (11.59)	23.0 (12.97)
	Min, Max	2.0, 65.5	1.4, 59.3	1.9, 58.0	1.8, 65.7
Most Common ^a Etiology of Epilepsy	%Other/Unknown	45	40	49	42
	%Idiopathic	16	18	13	20
	%Congenital/hereditary	8	15	15	13
	%Infection	9	13	7	12
	%Cranial Trauma	11	10	7	7
Seizure Frequency at Visit 2 ^b	N	100	96	100	97
	Mean (SD)	13.3 (14.03)	14.4 (18.65)	15.5 (15.27)	15.9 (16.31)
	Median	7.4	8.2	9.1	9.3
	Min, Max	3.2, 79.4	3.3, 127.7	3.4, 78.7	2.0, 87.9
Concomitant AED at Baseline	%Taking Any AED	100	100	100	100
	%Taking 1 AED	15	23	17	21
	%Taking 2 AEDs	76	71	73	69
	%Taking 3 or 4 AEDs	9	6	10	10
Most Common ^a Types of AED	%Carbamazepine	58	61	61	59
	%Valproic acid	26	13	28	21
	%Lamotrigine	24	22	17	22
	%Clobazam	16	20	21	12
	%Levetiracetam	16	16	13	20
	%Phenytoin	14	11	15	12
	%Phenobarbital	18	11	13	9
	%Topiramate	12	11	11	15

Abbreviations: SD=standard deviation; AED=antiepileptic drug; Min=minimum; Max=maximum

a Reported in ≥5% of the total ITT analysis set.

b Standardized to 4 weeks.

Source: Sponsor’s ISE Report Table 2.2.2-2.

3.1.2.3. Sponsor’s Efficacy Results

Primary efficacy results

Seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. Dunnett’s multiple comparison procedure was used for the comparison of the active treatment means to the placebo mean. Natural logarithm transformation was applied to standardised seizure $\text{Ln}(\text{standardized seizure frequency}+4)$. The standardized seizure frequency for a period is calculated by:

$$(\text{number of seizures} * 28 / \text{days in the period}).$$

The result was presented in Table 15:

Table 15. Study 302: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result)

Standardized Seizure Frequency per 4 weeks	Placebo (N=100)	Eslicarbazepine Acetate Dose Group		
		400 mg (N=96)	800 mg (N=100)	1200 mg (N=97)
N	99	95	88	85
LS Mean (back-transformed)	9.8	8.7	7.1	7.0
[95% CI on the Mean]	[8.7, 11.1]	[7.7, 9.9]	[6.2, 8.2]	[6.0, 8.1]
LSMean Difference to Placebo		-1.1	-2.7	-2.8
ANCOVA p-value ^a		0.423	0.002	0.001

a p-value from Dunnett's multiple comparison procedure for the comparison of the active treatment means to the placebo mean.

Source: Sponsor's *ISE* Report Table 2.2.3-1.

The ANCOVAs of reduction in seizure frequency per 4 weeks for the PP population during the 12-week maintenance period and for the ITT and PP populations during the 2-week titration + 12-week maintenance periods showed results similar to those obtained with the primary efficacy analysis.

The primary analysis was updated to incorporate several changes in deriving the efficacy variable. The updated p-value was consistent for the 800 mg group, but increased from 0.001 to 0.042 for the 1200 mg group (Table 16).

Table 16. Study 302: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
N	99	94	87	81
LSmean (SE)	10.0 (0.67)	9.2 (0.65)	7.6 (0.59)	8.0 (0.62)
95% CI	8.7, 11.3	7.9, 10.5	6.5, 8.8	6.8, 9.2
Log Difference in LSMean (SE)		-0.06 (0.061)	-0.18 (0.062)	-0.15 (0.063)
95% CI for Difference in LSMean		-0.20, 0.08	-0.33, -0.04	-0.30, 0.00
p-value		0.6524	0.0095	0.0420

Source: Sponsor's response to October 8, 2009 Request for information and is confirmed by FDA reviewer. Without imputation, baseline AED as covariate

The analysis with worst-case imputation of missing data, as described for Study 301, was still favorable. P-value changed from 0.0095 to 0.0314 for the 800 mg group and from 0.0420 to 0.0775.

Secondary efficacy results

The proportion of responders (defined as subjects with a $\geq 50\%$ reduction in seizure frequency from baseline) over the 12-week maintenance period was analysed similarly to Study 301. About one-third of patients in the ESL 800 mg and ESL 1200 mg groups were responders, compared to less than one-fifth of patients in the placebo and ESL 400 mg groups. The differences between the ESL 800 mg and ESL 1200 mg groups and the placebo group were statistically significant ($p=0.0114$ and 0.0122 , respectively, Table 17).

Table 17. Study 302: Responder Analysis (Sponsor updated result)

Responder	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
n/N (%) ^a	18/99 (18.2)	19/94 (20.2)	31/87 (35.6)	29/81(35.8)
Chi-square p-value ^b		0.8608	0.0114	0.0122
Odds Ratio		1.14	2.49	2.51
95% CI		0.56, 2.34	1.27, 4.88	1.27, 4.97

a. n/N=number of responders/number of subjects with seizure data in the maintenance period.

b. p-value from pairwise test of each active treatment group compared to placebo.

Source: Sponsor’s response to October 8, 2009 Request for information.

Covariate assessment and subgroup analyses

The following subgroup analyses and covariate assessment were from the Sponsor’s CSR and were not based on the updated datasets. Seizure frequency was compared between each active treatment group and the placebo group using an ANCOVA that models seizure frequency as a function of baseline seizure frequency, region, assigned treatment, and the treatment-by-region interaction. Additional analyses were performed as described above but variables such as age, race, and sex, carbamazepine use (yes/no), were used instead of region. Results showed that there were some significant interactions between treatment and region, race and carbamazepine use.

An ANCOVA that models seizure frequency as a function of baseline seizure frequency, assigned treatment, and number of concomitant AEDs was done. The number of concomitant AEDs did not appear to have an effect on seizure frequency.

Standardized seizure frequency was also assessed by seizure type, including simple partial, complex partial, partial evolving, and unclassified, using the same ANCOVA model as was used for the primary efficacy assessment.

Sponsor’s Efficacy Summary

ESL 800 mg and 1200 mg QD administered as adjunctive therapy to subjects with refractory simple or complex partial seizures resulted in statistically significantly lower standardized seizure frequency over a 12-week maintenance period compared to placebo. The percent reduction in seizure frequency over placebo was 16.5% and 13.9% for ESL 800mg/day and 1200mg/day, respectively. These results were supported by a statistically significantly higher rate of responders compared to placebo for these 2 dose groups. The ≥50% responder rates for placebo, ESL 800mg/day and 1200mg/day were 18%, 36% and 36%, respectively. Results were robust with consistent conclusions drawn from the analyses conducted on both the ITT and the PP analysis sets.

The percent reduction in partial seizure frequency over placebo was 5.8% (p-value = 0.6524) for ESL 400mg/day indicating a numerically greater but not statistically significant difference. The 50% responder rate was 20% (p-value = 0.8608) for ESL 400mg/day, which was similar to the placebo group.

Data from this trial in the sponsor’s opinion clearly demonstrate that ESL 800mg/day and 1200mg/day is an effective treatment.

3.1.2.4. Reviewer’s Results

Similar issues were indentified for this study except that reviews of seizure data were performed before study unblinding. The form of hard-codes was different from study 301.

Hard-codes

An Excel sheet named “SEIZURES.xls” was used to decide if a seizure was included in the analyses. It worked the same way as hard-codes on the seizure data. It was generated by the Sponsor and CRO based on blinded review on February 5th and 7th, 2007. The converted SAS file is located in [\\Cdsesub1\evsprod\NDA022416\0000\m5\datasets\bia-2093-302-part1\listings](#). Based on Sponsor’s response to September 8th, 2009 request for information, the flag variable “noseiz” was manually populated to seizure records that were determined as non-seizure by the review team. Out of a total of 53912 records, 2285(4.24%) was flagged as “not a seizure”. Of those flagged, 417 have a comment. The most common comment is that it is a few seconds during lunch or breakfast. In addition, during the review they indentified duplicate seizures, multiple seizures, cluster seizures and re-assessed the missing or implausible date and time. The reviewer conducted a sensitivity analysis removing the hardcodes. It showed that the result was not sensitive to the hardcodes.

Results

Descriptive Summary

Observed median relative reduction in seizure frequency for the ITT population during the 12-week maintenance period was 16.7%, 25.8%, 36.5% and 41.9% in the placebo group, ESL 400 mg, ESL 800 mg and ESL 1200 mg groups respectively.

Table 18. Study 302: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed)

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
Baseline	7.5	8.3	9.5	9.5
Titration	8.7	8.0	6.0	8.0
Maintenance	7.0	7.4	5.9	7.0
Percent Change from Baseline to Maintenance (%)	-5.6	-20.7	-32.6	-28.2

Source: FDA reviewer

Analyses of Primary Endpoint

The results showed that the 800 mg/day dose had statistically significantly lower seizure frequencies than placebo during assessment period. The 400mg and 1200 mg groups failed to show statistically significant result. Based on the analysis using non-conservative imputation on

ITT population, the percent reduction over placebo calculated by $100 \times (1 - \exp(\text{LSMean difference of the log standardized seizure frequency}))$ were 9.5%, 23.6% and 16.5% for 400mg, 800mg and 1200mg groups respectively.

Table 19. Study 302: FDA Analysis Results for the Primary Endpoint

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
Completers (with maintenance assessment)				
N	99	94	87	81
LSmean (SE)	9.0 (0.59)	8.3 (0.57)	6.6 (0.48)	7.0 (0.52)
95% CI	7.9, 10.2	7.3, 9.5	5.7, 7.6	6.0, 8.1
Log Difference in LSMean (SE)		-0.06 (0.085)	-0.27 (0.087)	-0.22 (0.089)
95% CI for Difference in LSMean		-0.26, 0.14	-0.48, -0.07	-0.43, -0.01
p-value		0.8031	0.0056	0.0354
ITT population (Conservative Imputation)				
N	100	96	98	94
LSmean (SE)	9.2 (0.62)	8.5 (0.58)	7.2 (0.50)	7.9 (0.55)
95% CI	8.1, 10.5	7.4, 9.7	6.3, 8.2	6.8, 9.0
Log Difference in LSMean (SE)		-0.07 (0.086)	-0.22 (0.086)	-0.14 (0.087)
95% CI for Difference in LSMean		-0.28, 0.13	-0.42, -0.02	-0.35, 0.06
p-value		0.7185	0.0276	0.2470
ITT population (Non-conservative Imputation)				
N	100	96	98	94
LSmean (SE)	9.2 (0.64)	8.2 (0.59)	6.8 (0.49)	7.5 (0.55)
95% CI	8.0, 10.5	7.2, 9.5	5.9, 7.8	6.5, 8.7
Log Difference in LSMean (SE)		-0.10 (0.089)	-0.27 (0.089)	-0.18 (0.090)
95% CI for Difference in LSMean		-0.31, 0.11	-0.48, -0.06	-0.39, 0.03
p-value		0.5368	0.0072	0.1143

Source: FDA reviewer

Site 395 was identified to have compliance issues by FDA inspection. Analyses excluding site 395 showed similar results.

Analyses of Secondary Endpoints

Percent of responder

The analyses of responders were conducted for ITT population of subject with post-baseline seizure information. Subjects who did not have seizure data during the maintenance period were treated in two ways:

- I. They were considered as non-responders, per sponsor SAP;
- II. Responses during titration were used.

P-values were computed from 2 tests:

- a. CMH test, per sponsor SAP;
- b. Continuity-Adjusted Chi-square test, using sponsor's ISE analysis method.

Table 20. Study 302: FDA Responder Analyses

Responder	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
I. Subjects w/o maintenance data as non-responder				
n/N (%) ^a	18/100 (18.0)	19/ 96 (19.8)	31/ 98 (31.6)	29/ 94 (30.9)
CMH p-value ^b		0.7493	0.0266	0.0373
Chi-square p-value ^b		0.8904	0.0396	0.0548
Odds Ratio		1.12	2.11	2.03
95% CI		0.55, 2.30	1.08, 4.10	1.04, 3.98
II. Impute Using Titration				
n/N (%) ^a	18/100 (18.0)	20/ 96 (20.8)	33/ 98 (33.7)	32/ 94 (34.0)
CMH p-value ^b		0.6169	0.0119	0.0109
Chi-square p-value ^b		0.7483	0.0183	0.0169
Odds Ratio		1.20	2.31	2.35
95% CI		0.59, 2.44	1.20, 4.48	1.21, 4.57

a. n/N=number of responders/number of subjects with seizure data in the maintenance period.

b. Unadjusted p-value from pairwise test of each active treatment group compared to placebo.

Source: FDA.

The analyses were slightly sensitive to handling of dropouts for the 1200 mg group. P-values varied from 0.0109 to 0.0548. The result of this secondary endpoint supported the primary analysis.

Percent change from baseline in seizure frequency

This was not one of the secondary endpoint specified in the study protocol. However, it was included in the submission. Analysis was conducted in a similar manner to the primary endpoint. Below was the result for ITT population with non-conservative imputation (using titration data for patients without maintenance data). Results indicate that all ESL dose groups showed greater but not statistically significant reductions from baseline compared to the placebo group in standardized seizure frequency. The 800 mg dose group had the highest reduction in seizure frequency (p-value 0.0521).

Table 21. Study 302: Percent Change from Baseline in seizure frequency

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
N	100	96	98	94
LSmean (SE) (%)	3.6 (6.47)	-10.8 (6.60)	-17.9 (6.53)	-5.3 (6.67)
95% CI (%)	-9.1, 16.4	-23.8, 2.2	-30.8, -5.1	-18.4, 7.9
Difference in LSMean (SE) (%)		-14.44 (9.242)	-21.57 (9.202)	-8.90 (9.302)
95% CI for difference in LSMean(%)		-36.25, 7.37	-43.29, 0.15	-30.85, 13.05
p-value		0.2773	0.0521	0.6574

Source: FDA

3.1.3. Supportive Study 303

3.1.3.1. Study Design and Endpoints

Study 303, titled “Efficacy and safety of BIA 2-093 as adjunctive therapy for refractory partial seizures in a double-blind, randomized, placebo-controlled, parallel-group, multicentre clinical study”, was initiated on 14 Dec 2004 and completed on 19 Jan 2007. The original protocol was issued on 09 Jul 2004 and amended 2 times. Amendment 1 was approved on 26 Nov 2004 to exclude all patients who were identified as legally or mentally handicapped at Visit 1 (screening). Amendment 2 was issued on 05 Jan 2007 to include the additional Mexico study sites. The Date of the SAP was 09 Jan 2007 and the final signature was collected on 12 Jan 2007.

This was a multicenter, randomized, double-blind, placebo-controlled study consisted of an 8-week baseline period, a 2-week titration, a 12-week maintenance period, and a 4-week tapering-off period. At the end of the baseline period, Patients were randomized for treatment with either ESL (1200 or 800 mg/day) or placebo during Part I. All patients were treated with ESL during Part II open-label extension phase.

The primary and secondary efficacy endpoints were the same as study 301.

Study 303 was included as supportive rather than for conclusive evidence of efficacy due to certain critical GCP deficiencies observed during clinical site audits in Mexico, where the majority of the subjects in this study were enrolled (details in sponsor module 2.5 P64).

3.1.3.2. Patient Disposition, Demographic and Baseline Characteristics

A total of 330 subjects were screened at 39 sites in 3 countries with the majority (209 patients) in Mexico, 51 patients in Spain, and 70 patients in Portugal. Of them 253 subjects were randomized into this study. A total of 195 patients completed Part I of the study. The pooled regions were as follows: Europe (Spain and Portugal), and Mexico

Table 22. Study 303: Subject Disposition

Disposition	Placebo	Eslicarbazepine Acetate Dose Group		Total
		800 mg	1200 mg	
Randomized	88	85	80	253
Completed the Study	66 (75.0)	70 (82.4)	59 (73.8)	195 (77.1)
Withdrew Prematurely	22 (25.0)	15 (17.6)	21 (26.3)	58 (22.9)
Withdrew during/end of Baseline	2 (2.3) ^a	0	0	2 (0.8)
Withdrew during/end of Titration	2 (2.3)	3 (3.5)	4 (5.0)	9 (3.6)
Withdrew during/end of Maintenance	11 (12.5)	9 (10.6)	15 (18.8)	35 (13.8)
Withdrew during/end of Taper	7 (8.0) ^b	3 (3.5)	2 (2.5)	12 (4.7) ^b
Reasons for Withdrawal ^c				
Unacceptable Adverse Event	7 (8.0)	8 (9.4)	9 (11.3)	24 (9.5)
Subject Non-compliance	2 (2.3)	3 (3.5)	5 (6.3)	10 (4.0)
Withdrawal of Consent	4 (4.5)	2 (2.4)	2 (2.5)	8 (3.2)
Lack of Efficacy	2 (2.3)	1 (1.2)	1 (1.3)	4 (1.6)
Protocol violation	2 (2.3)	0	0	2 (0.8)
Other	9 (10.2)	5 (5.9)	4 (5.0)	18 (7.1)

a One subject was not dosed and the other received a single dose prior to discontinuation for non-compliance.

b Includes 3 subjects who completed the tapering-off period but who did not continue into Part 2 of the study.

c A subject may have more than 1 reason for discontinuation.

Source: Sponsor's ISE Report Table 2.3.1-1.

The ITT data set consisted of all randomized patients with at least one administration of study medication and at least one post-baseline seizure frequency assessment. Of the 253 patients randomized, 7 subjects who did not have valid post-baseline efficacy assessments were excluded from the ITT data set. Note that in the ITT population, 39 subjects (16%, 16%, and 17% of subjects in the placebo, ESL 800 mg, and ESL 1200 mg groups, respectively) had baseline seizure frequency <4 per 4 weeks, a major protocol violation.

The demographic characteristics of the ITT analysis set were similar across the treatment groups. Race was classified as other for 62% of subjects and Caucasian for 38%. The high proportion of the 'other' category for race is related to reporting differences across countries. In Mexico, subjects of Hispanic ethnicity reported race as 'other' whereas in Portugal and Spain subjects of Hispanic ethnicity reported race as Caucasian.

Table 23. Study 303: Demographic and anthropometric characteristics (ITT Analysis Set)

Parameter	Statistic/Class	Placebo (N=84)	Eslicarbazepine Acetate Dose Group		Total (N=245)
			800 mg (N=84)	1200 mg (N=77)	
Age (years)	N	84	84	77	245
	Mean (SD)	38.0 (12.07)	36.8 (10.69)	35.4 (11.25)	36.8 (11.36)
	Min, Max	17.0, 77.0	18.0, 64.0	17.0, 68.0	17.0, 77.0
Sex	% Male	50	42	44	45
Race	% Caucasian	39	38	35	38
	% Other ^a	61	61	65	62
BMI (kg/m ²)	N	83	83	77	243
	Mean (SD)	25.7 (4.04)	25.5 (4.75)	26.9 (4.99)	26.0 (4.63)
	Min, Max	17.8, 36.5	15.4, 40.4	18.0, 44.0	15.4, 44.0

Abbreviations: SD=standard deviation; BMI=body mass index; Min=minimum; Max=maximum

a Subjects of Hispanic ethnicity from Mexico reported Race as 'other' whereas subjects from Portugal and Spain reported Race as Caucasian.

Reference: CSR 2093-303, Section 14, [Table 14.1-2.1.4](#) and [Table 14.1-2.2.4](#).

Source: Sponsor's ISE Report Table 2.3.2-1.

Median seizure frequency during the 8-week baseline period, standardized to 4-weeks, was 6.4, 7.7, and 6.0 seizures per 4 weeks in the placebo and ESL 800 mg and 1200 mg groups, respectively.

Table 24. Study 303: Baseline Disease Characteristics (ITT Analysis Set)

Parameter	Statistic/Class	Placebo (N=84)	Eslicarbazepine Acetate Dose Group	
			800 mg (N=84)	1200 mg (N=77)
Duration of Epilepsy (years)	N	84	84	77
	Mean (SD)	24.0 (13.19)	22.3 (11.78)	22.5 (12.91)
	Min, Max	1.1, 63.2	0.6, 51.1	1.0, 52.8
Most Common ^a Etiology of Epilepsy	%Other/Unknown	23	31	33
	%Idiopathic	42	42	38
	%Infection	8	10	14
	%Congenital	12	6	7
	%Cranial trauma	8	8	8
Seizure Frequency at Visit 2 ^b	N	84	84	77
	Mean (SD)	12.6 (17.93)	12.8 (18.23)	11.7 (12.25)
	Median	6.4	7.7	6.0
	Min, Max	0.0, 130.3	1.3, 150.2	0.0, 70.7
Concomitant AED at Baseline	%Taking Any AED	100	100	100
	%Taking 1 AED	19	25	14
	%Taking 2 AEDs	76	69	79
	%Taking 3 or 4 AEDs	5	6	6
Most Common ^a Types of AED	%Carbamazepine	69	50	47
	%Valproic acid	30	27	35
	%Levetiracetam	25	20	18
	%Phenytoin	12	21	17
	%Topiramate	11	17	19
	%Lamotrigine	13	11	16

Abbreviations: SD=standard deviation; AED=antiepileptic drug; Min=minimum; Max=maximum

a Reported in ≥5% of the total ITT analysis set.

b Standardized to 4 weeks.

Source: Sponsor's ISE Report Table 2.3.2-2.

3.1.3.3. Sponsor's Efficacy Results

Primary efficacy results

Seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. Dunnett's multiple comparison procedure was used for the comparison of the active treatment means to the placebo mean. The result is presented in Table 25:

Table 25. Study 303: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor CSR result)

Standardized Seizure Frequency per 4 weeks	Placebo (N=84)	Eslicarbazepine Acetate Dose Group	
		800 mg (N=84)	1200 mg (N=77)
N	79	80	69
LS Mean (back-transformed)	7.3	5.7	5.5
[95% CI on the Mean]	[6.3, 8.5]	[4.9, 6.7]	[4.6, 6.5]
LSMean Difference to Placebo		-1.6	-1.9
ANCOVA p-value ^a		0.048	0.021

a p-value from Dunnett's multiple comparison procedure for the comparison of the active treatment means to the placebo mean.

Source: Sponsor's ISE Report Table 2.3.3-1.

The primary analysis was updated to incorporate changes in deriving the efficacy variable. The p-value was increased from 0.048 to 0.1476 for the 800 mg group, and from 0.021 to 0.0455 for the 1200 mg group. Standardized seizure frequency over the 12-week maintenance period was statistically significantly lower in 1200 mg dose groups compared to the placebo group, and was numerically lower but not statistically significant for the 800 mg group (Table 26).

Table 26. Study 303: Seizure Frequency per 4 Weeks over the Maintenance Period (Sponsor updated result)

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
N	79	---	80	70
LSmean (SE)	7.9 (0.96)		6.5 (0.82)	6.1 (0.82)
95% CI	6.1, 9.9		5.0, 8.3	4.6, 7.8
Log Difference in LSMean (SE)			-0.12 (0.069)	-0.16 (0.072)
95% CI for Difference in LSMean			-0.28, 0.03	-0.32, 0.00
p-value			0.1476	0.0455

Source: Sponsor's response to October 8, 2009 Request for information and is confirmed by FDA reviewer. Without imputation, baseline AED as covariate

Based on the CSR, for the 12-week maintenance period, the proportion of responders in the ITT analysis set was 22.6% in the placebo group compared to 34.5% and 37.7% in the ESL 800 mg and 1200 mg groups, respectively. The comparison of results between the ESL 1200 mg group and the placebo group was statistically significant (p=0.020); the comparison of the ESL 800 mg group to the placebo group was not significant (p=0.106). Results were similar in the PP analysis set.

The blind was broken for 5 patients from site 709 immediately after study completion. To assess the potential impact this deviation had on results from this study, a restricted ITT population, in which data for these 5 patients were removed from the ITT population, was used for efficacy analyses. For the primary efficacy analysis, the p-values are 0.019 and 0.077 respectively for differences between ESL 800 mg and ESL 1200 mg groups and the placebo group in standardized seizure frequency. Results were similar for key secondary efficacy analyses.

Subgroup/covariate analyses

An exploratory ANCOVA was performed with a model that included the number of concomitant AEDs as a factor in addition to treatment over the 12-week maintenance period. The number of concomitant AEDs did not appear to have an effect on seizure frequency for the ITT population. In the PP population, however, the number of concomitant AEDs did have a significant effect on seizure frequency during all time periods analyzed ($p \leq 0.0192$).

The median relative reduction in frequency of simple partial seizures during the 12-week maintenance period was highest in the ESL 1200 mg group (61.9%), compared with 44.4% in the placebo group and 42.4% in the ESL 800 mg group. However, in analyses including the 2-week titration or 2-week titration and 4-week tapering-off periods, median relative reduction in frequency was less than the placebo group in both active treatment groups.

The median relative reduction in the frequency of complex partial seizures occurred in a dose dependent manner during the 12-week maintenance period, with 30.8% in the placebo group, 51.7% in the ESL 800 mg group, and 62.7% in the ESL 1200 mg group. Results are similar for other study periods.

3.1.3.4. Reviewer's Results

See Section 3.1.2.4 for issues identified for this study. The reviewer used the same approach as previous studies to analyze the data.

Descriptive Summary

Observed median relative reduction in seizure frequency for the ITT population during the 12-week maintenance period was 15.2%, 36.4% and 43.1% in the placebo group, ESL 800 mg and ESL 1200 mg groups respectively (Table 27).

Table 27. Study 303: Median Seizure frequency during Baseline, Titration and Maintenance Phase (Observed)

	Placebo	Eslicarbazepine Acetate Dose Group		
		-	800 mg	1200 mg
Baseline	6.5	-	7.8	6.1
Titration	8.0	-	6.4	7.0
Maintenance	5.8	-	5.0	4.0
Percent Change from Baseline to Maintenance (%)	-15.2	-	-36.4	-43.1

Source: FDA

Analyses of Primary Endpoint

The results showed that the 1200 mg/day dose had statistically significantly lower seizure frequencies than placebo during assessment period. The 800 mg groups failed to show statistically significant result (Table 28). Based on the analysis using non-conservative imputation on ITT population, the percent reduction over placebo calculated by $100*(1-$

exp(LSmean difference of the log standardize seizure frequency)) were 18.9% and 22.8% 800mg and 1200mg groups respectively.

Table 28. Study 303: FDA Analysis Results for the Primary Endpoint

	Placebo	Eslicarbazepine Acetate Dose Group		
		-	800 mg	1200 mg
Completers (with maintenance assessment)				
N	79	-	80	70
LSmean (SE)	6.7 (0.59)	-	5.2 (0.47)	4.9 (0.48)
95% CI	5.6, 8.0	-	4.4, 6.2	4.0, 5.9
Log Difference in LSMean (SE)		-	-0.21 (0.107)	-0.27 (0.111)
95% CI for Difference in LSMean		-	-0.45, 0.03	-0.51, -0.02
p-value		-	0.0882	0.0322
ITT population (Conservative Imputation)				
N	84	-	84	77
LSmean (SE)	7.0 (0.59)	-	5.4 (0.47)	5.3 (0.48)
95% CI	5.9, 8.3	-	4.6, 6.4	4.4, 6.3
Log Difference in LSMean (SE)		-	-0.22 (0.104)	-0.25 (0.106)
95% CI for Difference in LSMean		-	-0.45, 0.01	-0.49, -0.01
p-value		-	0.0594	0.0363
ITT population (Non-conservative Imputation)				
N	84	-	84	77
LSmean (SE)	6.8 (0.58)	-	5.3 (0.47)	5.0 (0.47)
95% CI	5.7, 8.0	-	4.4, 6.3	4.1, 6.0
Log Difference in LSMean (SE)		-	-0.21 (0.106)	-0.26 (0.108)
95% CI for Difference in LSMean		-	-0.45, 0.03	-0.50, -0.02
p-value		-	0.0887	0.0335

Source: FDA

Analyses of Secondary Endpoints

Percent of responder

The analyses of responders were conducted for ITT population of subject with post-baseline seizure information. Subjects who did not have seizure data during the maintenance period were treated in two ways:

- I. They were considered as non-responders, per sponsor SAP;
- II. Responses during titration were used.

P-values were computed from 2 tests:

- a. CMH test, per sponsor SAP;
- b. Continuity-Adjusted Chi-square test, using sponsor's ISE analysis method.

The result of this secondary endpoint supported the primary analysis (Table 29).

Table 29. Study 303: FDA Responder Analysis

Responder	Placebo	Eslicarbazepine Acetate Dose Group		
		-	800 mg	1200 mg
I. Subjects w/o maintenance data as non-responder				
n/N (%) ^a	19/ 84 (22.6)	-	28/ 84 (33.3)	31/ 77 (40.3)
CMH p-value ^b		-	0.1230	0.0160
Chi-square p-value ^b		-	0.1691	0.0247
Odds Ratio		-	1.71	2.31
95% CI		-	0.86, 3.39	1.16, 4.57
II. Impute Using Titration				
n/N (%) ^a	21/ 84 (25.0)	-	29/ 84 (34.5)	34/ 77 (44.2)
CMH p-value ^b		-	0.1783	0.0107
Chi-square p-value ^b		-	0.2375	0.0167
Odds Ratio		-	1.58	2.37
95% CI		-	0.81, 3.09	1.22, 4.63

a. n/N=number of responders/number of subjects with seizure data in the maintenance period.

b. Unadjusted p-value from pairwise test of each active treatment group compared to placebo.

Source: FDA.

Percent change from baseline in seizure frequency

This was not one of the secondary endpoint specified in the study protocol. However, it was included in the submission. Analysis was conducted in a similar manner to the primary endpoint. Below was the result for ITT population with non-conservative imputation (using titration data for patients without maintenance data). Results indicate that the difference between ESL dose groups and placebo are not statistically significant.

Table 30. Study 303: Percent Change from Baseline in seizure frequency

	Placebo	Eslicarbazepine Acetate Dose Group		
		-	800 mg	1200 mg
N	84	-	84	77
LSmean (SE) (%)	-2.0 (9.30)	-	-19.3 (9.30)	-18.8 (9.71)
95% CI (%)	-20.3, 16.3	-	-37.6, -1.0	-37.9, 0.3
Difference in LSMean (SE) (%)		-	-17.35 (13.146)	-16.84 (13.444)
95% CI for Difference in LSMean (%)		-	-46.62, 11.92	-46.77, 13.09
p-value		-	0.3165	0.3522

Source: FDA

3.2. Evaluation of Safety

Please see the clinical review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1. Gender, Race and Age

4.1.1. Gender

About 51% of all patients were female and 49% were male in both study 301 and 302. Overall, there was no compelling evidence that the treatment effect varied by gender ($p=0.54$). From the table below, the drug is effective in both genders for 800 and 1200 mg/day groups. The drug effect was numerically larger in the females than the males. For the females, the drug effect was numerically larger in 1200 mg/day group; for the males, the drug effect was numerically larger in 800 mg/day group. The interaction between gender and treatment was not significant ($p=0.72$).

Table 31. Treatment Effect by Gender in Pooled Population from Study 301 and Study 302

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
Female				
N	102	106	92	96
LSmean (SE)	8.6 (0.62)	7.1 (0.51)	6.2 (0.49)	6.2 (0.48)
95% CI	7.5, 9.9	6.2, 8.2	5.3, 7.3	5.3, 7.2
Log Difference in LSMean (SE)		-0.17 (0.090)	-0.28 (0.094)	-0.29 (0.093)
95% CI for Difference in LSMean		-0.38, 0.05	-0.51, -0.06	-0.51, -0.08
p-value		0.1613	0.0071	0.0047
Male				
N	100	88	104	95
LSmean (SE)	7.4 (0.49)	7.2 (0.51)	5.7 (0.38)	5.8 (0.41)
95% CI	6.5, 8.4	6.2, 8.2	5.0, 6.5	5.1, 6.7
Log Difference in LSMean (SE)		-0.03 (0.085)	-0.23 (0.082)	-0.21 (0.084)
95% CI for Difference in LSMean		-0.23, 0.17	-0.42, -0.04	-0.41, -0.01
p-value		0.9775	0.0150	0.0363

4.1.2. Race

About 94% of patients were Caucasian. All patients in study 301 were Caucasian. The next largest group was Black with 23 patients (2.9%). Based on the limited data for non-Caucasian races, there was no compelling evidence that the treatment effect varied significantly with race ($p=0.96$ comparing treatment effects for Caucasians to treatment effects for others).

4.1.3. Age

The range of age is 18 to 75, and the median age was 37. There were only 30 patients age 60 or above, so no reliable analysis of efficacy can be done in this subgroup. A test for a differential effect according to age was not significant ($p=0.12$) in the pool of studies 301 and 302. This test assumed seizure rates were linear in age, allowing for a separate linear relationship for each

group. It concluded that the slopes were not significantly different implying that there was insufficient evidence to conclude that the treatment effect varied significantly with age.

4.2. Other Special/Subgroup Populations

4.2.1. Region

4.2.1.1. Study 301

In the 1200mg group, Eastern Region had the highest response rate and percent reduction, while the Central Region had the smallest response, which was similar to the placebo group. In the 800mg group, Eastern and Central Region responded similarly. Placebo group in Central Region had numerically larger response rate. Central Region and Eastern Region showed different dose-response relationship, mostly due to the large difference in the high dose group. There were not enough subjects in the Western Region. Region and treatment-by-region interaction did not have significant effect on the seizure frequency. The p-values were 0.23 and 0.58 respectively.

Table 32. Treatment Effect by Region in Study 301

	Eslicarbazepine Acetate Dose Group			
	Placebo	400 mg	800 mg	1200 mg
Central Region				
N	51	51	50	49
Median Percent Change	-23.23	-26.85	-36.63	-24.29
Response Rate	25.49	21.57	36.00	30.61
Eastern Region				
N	44	41	41	41
Median Percent Change	-15.43	-26.43	-35.63	-55.40
Response Rate	13.64	29.27	34.15	60.98
Western Region				
N	7	7	7	8
Median Percent Change	-15.04	3.40	3.37	-40.62
Response Rate	14.29	14.29	28.57	37.50

Source: FDA

4.2.1.2. Study 302

The highest response rate was in Australia and South Africa Region in the 800mg group. Australia and South Africa Region had higher placebo effect. Region and treatment-by-region interaction did not have significant effect on the seizure frequency. The p-values were 0.15 and 0.72 respectively.

Table 33. Treatment Effect by Region in Study 302

	Placebo	Eslicarbazepine Acetate Dose Group		
		400 mg	800 mg	1200 mg
Australia and South Africa				
N	18	15	18	15
Median Percent Change	-15.7	-23.1	-58.8	-20.8
Response Rate	11.1	33.3	55.6	46.7
Europe				
N	27	28	29	30
Median Percent Change	-1.8	-21.9	-31.3	-16.5
Response Rate	11.1	17.9	31.0	23.3
South America				
N	55	53	53	52
Median Percent Change	-3.3	-19.9	-24.1	-29.2
Response Rate	23.6	18.9	30.2	40.4

Source: FDA

4.2.2. Carbamazepine Use

In the pool of studies 301 and 302, Carbamazepine use and treatment-by-carbamazepine use interaction did not have significant effect on the seizure frequency. The p-values were 0.17 and 0.87 respectively.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Collective Evidence

Below are the results for ITT population at maintenance period (titration period data was used if no maintenance data). The two pivotal studies supported the efficacy for 800 mg/day dose. The efficacy result for 1200 mg/day dose had large variation among the studies. All studies suggested a lack of significant difference between the 400 mg dose and placebo.

Table 34 Primary endpoint: maintenance seizure frequency

		Placebo	Eslicarbazepine Acetate Dose Group		
			400 mg	800 mg	1200 mg
301	LSmean	6.9	6.2	5.2	4.8
	p-value		0.5136	0.0125	0.0007
302	LSmean	9.2	8.2	6.8	7.5
	p-value		0.5368	0.0072	0.1143
303	LSmean	6.8		5.3	5.0
	p-value			0.0887	0.0335

Table 35 Secondary endpoint: percent of responder during maintenance

		Placebo	Eslicarbazepine Acetate Dose Group		
			400 mg	800 mg	1200 mg
301	n/N (%)a	20/102 (19.6)	23/ 98 (23.5)	34/ 98 (34.7)	42/ 97 (43.3)
	Chi-square p-value		0.6225	0.0249	0.0006
302	n/N (%)a	18/100 (18.0)	20/ 96 (20.8)	33/ 98 (33.7)	32/ 94 (34.0)
	Chi-square p-value		0.7483	0.0183	0.0169
303	n/N (%)a	21/ 84 (25.0)		29/ 84 (34.5)	34/ 77 (44.2)
	Chi-square p-value			0.2375	0.0167

Unadjusted p-value from pairwise test of each active treatment group compared to placebo.

Table 36 Percent reduction from baseline

		Placebo	Eslicarbazepine Acetate Dose Group		
			400 mg	800 mg	1200 mg
301	LSmean	-7.7	-15.9	-28.4	-29.6
	p-value		0.6391	0.0373	0.0262
302	LSmean	3.6	-10.8	-17.9	-5.3
	p-value		0.2773	0.0521	0.6574
303	LSmean	-2.0		-19.3	-18.8
	p-value			0.3165	0.3522

Study Design Issue

In this NDA, the trial participants were instructed to update seizure diary only when they experienced a seizure. As a result, “0” seizure was not recorded. Therefore, true zero seizures could not be differentiated from missing seizure data (patient did not record a seizure, missed a visit, or did not return diary card, etc.), and “no seizure data” was assumed as no occurrence of

seizure events in the analysis. A worst-case-scenario analysis assessed the effect of the part of missing data that were caused by unreturned diary cards. This analysis showed favorable results, although lost significance for the 800 mg/day dose in one of the study.

Study Conduct Issues

The sponsor has performed extensive hardcoding in the program that essentially changed the values of the variables in the database. In Study 301, 559 hard-codes were used when the sponsor generated the analysis datasets. Some hard-codes were generated from unblinded review of seizure data after data lock. In Study 302, a manually generated file was used to modify data. Upon reviewing the hardcodes, it seemed that the hardcodes were generally attempts to correct errors in the database. A sensitivity analysis with the removal of all the hardcodes associated with seizure data indicated that the result was insensitive to the hardcodes.

Hardcoding over rides the database controls in the clinical data management systems and may compromise study data integrity. Although the sponsor did not seem to intentionally mislead FDA by manipulating data, the extensive hardcoding however, indicated that the study was not well conducted and the data quality/reliability was questionable.

Statistical Analysis Issues

The sponsor specified the ITT population for the efficacy analysis as all randomised patients with at least one administration of study medication and at least one post-baseline (titration and /or maintenance period) seizure frequency assessment. However, the SAP also specified that missing values were not imputed. The majority of efficacy evaluations were performed for the maintenance period; therefore, subjects who discontinued the study during the titration period were not included in the primary and key secondary analyses. As more subjects in high dose group withdrew in early treatment phase, this analysis may be biased. The results were robust to the handling to the early drop-outs for the 800 mg/day group, but a little sensitive for the 1200 mg/day dose group as evidenced by the loss of significance in one study when a patient's missing data after dropout during the titration period was imputed.

Natural logarithm transformation was applied to standardised seizure frequency (denoted as S) $\ln(S+4)$. The purpose of logarithm transformation was to approximate the normal distribution, and the positive number added before the logarithm transformation was to avoid log of zero. No justification was given in the SAP for using this transformation instead of ' $\ln(S+1)$ '. However, the $\ln(S+1)$ approximates the normal distribution better than $\ln(S+4)$, as confirmed by Goodness-of-Fit Tests for Normal Distribution. In addition, the statistical inference was based on LS Mean difference of the log standardize seizure frequency, which translates to the percent reduction over placebo by $100 \times (1 - \exp(\text{LS Mean difference of the log standardize seizure frequency}))$. This is approximately $(\text{Treatment group frequency} - \text{Placebo group frequency}) / (\text{Placebo group frequency} + 4)$. When the constant used in the logarithm transformation is large, the percent reduction over placebo is underestimated. To better approximate the normal distribution and to estimate the treatment effect, the reviewer presented results obtained using transformation ' $\ln(S+1)$ '. The change to the transformation does not drastically affect the results.

SAP stated that seizure frequency was compared among the treatment groups by using an analysis of covariance (ANCOVA) that models seizure frequency as a function of baseline seizure frequency and treatment. In the updated analyses by sponsor, an additional term (the number of AEDs used at baseline) was included in the model. No justification of including this covariate was given, and this covariate did not seem to have a significant effect on the seizure frequency. The reviewer presented the results from ANCOVA models that did not include the additional term of the number of AEDs. The analysis results had minimal changes with or without this term.

5.2 Conclusions and Recommendations

The data seem to support the efficacy of ESL as adjunctive therapy to subjects with refractory simple or complex partial seizures. In both pivotal studies, the 800 mg/day dose resulted in statistically significantly lower standardized seizure frequency over a 12-week maintenance period compared to placebo. In the supportive phase III study, the estimate of the treatment effect for the 800 mg/day group was similar to those in the pivotal studies, although the p-value was slightly larger than 0.05, possibly due to the smaller ITT analysis set and larger variance.

The 1200 mg/day dose was significantly better than placebo in one of the two pivotal studies. In the other study, the results were sensitive to the handling of the dropouts. In the analysis of the ITT population with non-conservative imputation for the early drop-outs, the p-value was 0.1143. In the supportive phase III study, the 1200 mg/day dose was marginally significant.

There is no compelling evidence that the 1200 mg/day dose provided added improvement over the 800 mg/day dose. In study 301, the 1200 mg dose group showed incremental efficacy over the 800 mg group in standardized seizure frequencies and the proportion of responders. However, the additional efficacy in the 1200 mg dose group was *not* demonstrated in study 302, where the 800 mg group had lower standardized seizure frequencies during the maintenance period.

The 400 mg/day dose was not significant in both pivotal studies.

APPENDICES

Appendix I. Statistical Questions to Sponsor on September 8 2009

Question 1

As zeros are not recorded on diary cards and in the dataset, please explain how ‘missing seizure records’ (for example, a diary card not returned for a visit) and ‘no occurrence of seizure’ are differentiated. If they cannot be differentiated, please assess its impact on the efficacy results.

Question 2

Please explain that for subjects who withdrew early, how the end dates were determined for the purpose of calculating the number of days and deriving the seizure frequency.

Question 3

For study 302, you indicated that “Bial and its partner [REDACTED] ^{(b) (4)} (CRO) held meetings to identify records which are not regarded as seizures”. Please clarify the definition of seizure and how the flag variable “noseiz” was generated. Please submit evidence that this was done before the study was unblinded.

Question 4

The disposition table for study 301 indicates that only 82 subjects entered maintenance period in 1200 mg group. However, the primary efficacy result shows that 94 subjects had efficacy measures during maintenance period (without imputation). Please explain the discrepancy.

Question 5

Explain the discrepancies in the disposition tables from individual studies and ISE. For example, based on the tables of individual studies, 87 subjects (33 and 54 from study 301 and 302 respectively) withdrew due to unacceptable adverse event based on the individual table. However, the pooled analysis shows that only 73 subjects withdrew due to AE.

Appendix II. Statistical Questions to Sponsor on October 8 2009

Question 1

The sponsor has performed a large degree of hardcoding in the program that essentially changed the values of the variables in the database. For example, the dataset creation program for study 2093_301 ([sco_bia_2093_301_derived_data_part_i.pdf](#)) has 193 places with "if patid=xxx then variable ABC=zzz" indicating a value for a specific subject is changed in the program. Hardcoding is generally discouraged because 1) too much hard coding over rides the database controls in the clinical data management systems and may compromise study data integrity and 2) other people can not reproduce the results based on the raw datasets without knowing all the data changes.

We request the following:

- 1) Please, explain when and how the changed values are determined.
- 2) Please submit one dataset (or a few datasets if necessary) with original variables and a program to generate the final efficacy variables (seizure frequency during maintenance period, and seizure frequency during titration and maintenance period) based on this dataset. The dataset should only include those raw variables (not derived variables) that are needed to generate the efficacy variables. Use the last diary return date as the end date of early terminated subjects.
- 3) Please list the source (CRF, team meeting, etc) of each variable in this dataset, and the hardcodes used in the program. Give explanations/source documents of the hardcodes when necessary.
- 4) Rerun the primary efficacy analysis using the newly derived efficacy variable for each of the phase 3 studies.

Question 2

The period for the purpose of counting the number of days does not seem to be consistent with the period for counting the number of seizures for early terminated subjects in ISE. For example, for subject 90040 (study 301/1200mg group) the 5 seizures after 22DEC04 were not counted as maintenance seizures for the reason of EOT. However, the number of days after 22DEC04 was counted as part of the maintenance period, resulting in the dilution of the standardized seizure to 1.7.

Please explain or correct when deriving the new efficacy variables mentioned in Question 1.

Question 3

You mentioned in your reply to Biostatistics question 3 on September 8 that duplicate seizures are identified in the blinded data review meetings. This process, however, does not seem to be followed in the ISE. For example, there seems to be 128 duplicate seizures counted for the maintenance period for subject 2093301-172-90435.

Make sure duplicate seizures are handled correctly when deriving the new efficacy variables mentioned in Question 1.

Question 4

As handling of the cluster seizures is not pre-specified in the SAP, please conduct a sensitivity analysis in which each individual seizure is counted instead of each cluster group.

Appendix III. Statistical Questions to Sponsor on November 4 2009.

We have received your response to FDA Sept. 8 Biostatistics Question 1. In your reply you indicated that the studies were to collect all diaries (used or un-used). Please provide a summary of patient diary compliance (e.g., percentage of diaries returned) and assess the impact of non-compliance on efficacy results. For each phase 3 study, please provide us the following analysis results, derived datasets for these analyses, listing of involved raw datasets, the involved raw datasets if they are not already submitted, and programs.

1. Summary of Patient Diary Compliance. The following table is an example and the sponsor can provide other tables such as summaries by study visit/phase if deemed necessary.

Percentage of Diaries Returned	Placebo n(%)	ESL 400 mg n(%)	ESL 800 mg n(%)	ESL 1200 mg n(%)
All patients				
100%				
80-100%				
60-80%				
<60%				

2. Analysis of the primary efficacy endpoint for subsets of ITT patients with 100% diary compliance and patients with >80% diaries returned.
3. Please provide a worst case analysis. The worst case scenario analysis may include an analysis that assumes rates for times that the diaries were missing as follows:

Missing diary during the experimental period:

- a) On placebo - assume missing diary rates are 0.
- b) On drug- assume missing diary rates are equivalent to the baseline rate.

Missing Diary during the baseline:

- a) On placebo- assume rates equivalent to rate during experimental period
- b) On drug- assume rates are 0.

For periods with no single diary card returned and no seizure reported (refer to table 1 in your response to Sept. 8 Biostatistics Question 1), exclude such periods from all the efficacy analyses requested.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22416	ORIG-1	SEPRACOR INC	SEP-0002093 ESLICARBAZEPINE ACETATE

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

/s/

XIANG LING
03/16/2010

KUN JIN
03/16/2010
I concur with this review.

KOOROS MAHJOOB
03/16/2010
I read the review and discussed my views with the primary reviewer.
My editorial and views are incorporated in this version and I concur with it.



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

IND Number: 22,416 / Serial 000

Drug Name: Eslicarbazepine acetate

Indication: Treatment of epilepsy

Applicant: Sepracor Inc.

Date: Submitted 29 March 2009

Review Priority: Standard

Biometrics Division: Division 6

Statistical Reviewer: Steve Thomson

Concurring Reviewer: Team Leader: Karl Lin, Ph. D.

Medical Division: Neurology Products

Toxicologist: Reviewer: Christopher Toscano, Ph.D.
Lois Freed, Ph.D.

Project Manager: Dorothy Demczar, PharmD.

Keywords: Bayesian analysis, 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	5
<i>Study FLG0024: 104 Week Oral (Gavage) Carcinogenicity Study in the Mouse</i>	5
1.3. STATISTICAL ISSUES AND FINDINGS	5
1.3.1. <i>Statistical Issues</i>	5
1.3.2. <i>Statistical Findings</i>	9
2. INTRODUCTION	10
2.1. OVERVIEW	10
2.2. DATA SOURCES	10
3. STATISTICAL EVALUATION	10
3.1. EVALUATION OF EFFICACY	10
3.2. EVALUATION OF SAFETY	10
3.2.1. <i>Study FLG0024: 104 Week Oral (Gavage) Carcinogenicity Study in the Mouse</i>	10
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS.....	17
5. SUMMARY AND CONCLUSIONS.....	18
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE.....	18
5.2. CONCLUSIONS AND RECOMMENDATIONS	18
APPENDICES:	19
APPENDIX 1. SURVIVAL ANALYSIS.....	19
APPENDIX 2. BAYESIAN ANALYSIS OF SURVIVAL.....	22
APPENDIX 3. POLY-K TUMORIGENICITY ANALYSIS	27
APPENDIX 4. BAYESIAN TUMORIGENICITY ANALYSIS	31
APPENDIX 5. REFERENCES.....	34

1. EXECUTIVE SUMMARY

According to the Sponsor the objective of this study was to assess the influence of eslicarbazepine acetate, compound BIA 2-093, on tumour formation when administered orally to mice, by gavage for up to 104 weeks. The original Sponsor was Bial in Portugal. This study was conducted by (b) (4)

1.1. Conclusions and Recommendations

For this study 54 CrI:CD-1 (ICR) BR VAF/PLUS derived strain mice of each gender were allocated to four different treatment groups: a vehicle control (0.5% hydroxypropylmethylcellulose) and three increasing doses of compound BIA 2-093, Eslicarbazepine acetate, at 100, 250, or 600 mg/kg/day, respectively. For the first week the 600 mg/kg/day group was actually dosed at 250 mg/kg/day. Animals were dosed once daily by oral gavage, 7 days a week for at least 103 weeks for males and 104 weeks for females until the day before necropsy.

The statistical significances of the tests of differences in survival across treatment groups are given in Table 1 below. Tests of homogeneity over all groups, dose related trend and the pairwise differences between the high dose group and the vehicle control were performed. Kaplan Meier survival curves for survival as a function of dose are provided in Appendix 1. In female mice, these Kaplan-Meier curves are generally intertwined, with no strong evidence of trend or lack of homogeneity over all dose groups (all $p \geq 0.3259$). In male mice survival in the high dose group is lower than the other groups, particularly early in the study, resulting in the statistically significant tests of the null hypothesis of homogeneity (logrank $p=0.0215$, Wilcoxon $p=0.0189$). However, this separation in survival is generally earlier in the study. By the end of the study survival in the vehicle control and the Eslicarbazepine high dose group are fairly close, which explains the statistically non-significant result comparing the high dose group to the control (both $p \geq 0.7171$). The wandering of the survival curves in male mice seems to be the reason for the somewhat ambiguous results of the simple test of trend (Logrank $p=0.1630$, Wilcoxon $p=0.0572$). As discussed in Appendix 1, there is equivocal evidence of a statistically significant departure from simple trend in male mice, but no corresponding evidence in female mice.

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

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Homogeneity over Groups 1-4	0.0215	0.0189	0.5682	0.3259
Trend over Groups 1-4	0.1630	0.0572	0.5499	0.6809
Comparison of High and Low	0.7171	0.7474	0.8481	0.6389

Appendix 2 presents an experimental Bayesian analysis of dose related survival. This analysis does not assume proportional hazards as would be assumed in an analysis based on the

Cox model. This analysis suggests that in both genders, differences among treatment groups are primarily due to early mortality in the high dose group, particularly in males.

Table 2 below lists tumors that are potentially statistically significant using the so-called poly-k modification of the Cochran-Armitage test (please see Section 1.3.1.3). Complete incidence tables are given in Appendix 3. Using the incidence in the vehicle control group to define the rarity of the neoplasms, in male all of these tumors are classified as common (incidence > 1%) while the corresponding tumors in females would be classified as rare (incidence ≤ 1%). Applying the Haseman-Lin-Rahman rules, as discussed in section 1.3.1.4 below, for a single species study, rare tumors could be considered statistically significant if the observed p-value is 0.05 or less, while common tumors would be considered as statistically significant if the observed p-value is 0.01 or less. In male mice the test of trend over dose in hepatocellular adenoma was statistically significant ($p=0.0086 < 0.01$), as was the test in hepatocellular carcinoma and pooled adenoma and carcinoma ($p < 0.00005 < 0.01$). In female mice, results were even more highly statistically significant, for all three tests of trend in adenoma, carcinoma, and pooled adenoma and carcinoma (all $p < 0.00005 < 0.05$). In male mice the test comparing the high dose group to the vehicle control group in adenomas was not quite statistically significant ($p = 0.0171 > 0.01$). However, in both genders, the tests comparing the high dose group to controls were highly statistically significant in both carcinomas and pooled adenomas and carcinomas ($p = 0.00005 < 0.01, 0.05$). The test comparing the high dose group to controls in adenomas in female mice was also highly statistically significant ($p = 0.0001 < 0.05$). In male rats the pairwise test of the medium dose group to the vehicle controls in adenoma was also statistically significant ($p = 0.0078 < 0.01$), as was the pairwise test in carcinoma and pooled adenoma and carcinoma (both $p < 0.00005 < 0.01$). No other pairwise comparison was even statistically significant at a nominal 0.05 level.

Table 2. Potentially Statistically Significant Neoplasms in Mice

Tumor	Veh	Low	Med	High	Trend	Veh vs Low	Veh vs Med	Veh vs High
Male Mice								
Liver								
ADENOMA - BENIGN	12	18	26	23	0.0086	0.2653	0.0078	0.0171
HEPATOCELLULAR CARCINOMA	5	10	29	32	0.0000	0.2848	0.0000	0.0000
Adenoma/Hepato. Carc.	15	24	38	40	0.0000	0.1784	0.0000	0.0000
Female Mice								
Liver								
ADENOMA - BENIGN	0	2	1	12	0.0000	0.2680	0.5143	0.0001
HEPATOCELLULAR CARCINOMA	0	0	2	15	0.0000	1.0000	0.2609	0.0000
Adenoma/Hepato. Carc.	0	2	2	24	0.0000	0.2680	0.2609	0.0000

Note that Appendix 4 has an experimental Bayesian assessment of the incidence of neoplasms. The results in that appendix are consistent with assessments above for each gender, though some analysts might prefer to indicate that the results above are consistent with the analysis in that appendix.

1.2. Brief Overview of the Studies

This submission had one mouse study:

Study FLG0024: 104 Week Oral (Gavage) Carcinogenicity Study in the Mouse

For this study four groups of 54 Crl:CD-1 (ICR) BR VAF/PLUS derived strain mice of each gender were allocated to four different treatments: a vehicle control (0.5% hydroxypropylmethyl-cellulose) and doses of compound BIA 2-093, Eslicarbazepine acetate, at 100, 250, or 600 (mg/kg/day). Note that the nominal high dose group, the 600 mg/kg/day group was actually dosed at 250 mg/kg/day for the first week of the study “in an attempt to alleviate the severity of the clinical signs” (page 14 of report). Animals were dosed once daily by oral gavage, 7 days a week for at least 103 weeks for males and 104 weeks for females until the day before necropsy.

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 details of the survival analyses, tests on tumorigenicity, multiplicity of tests on neoplasms, and the validity of the designs.

1.3.1.1. Animal Husbandry and Its Effect on Analysis:

The Sponsor reports that: “Males were housed in groups of 3 and females were housed in groups of 5, 4 and 3 ...” (page 24 of report). Effects of competition within a cage can be expected to induce negative correlations in many within cage responses. Propinquity could induce positive correlations. Tests based on the simple animal within cage scores will invariably treat the responses as being stochastically independent across animals, which would imply that variances are likely to be over- or under- estimated and thus that tests of differences based on these scores would tend to be conservative or anti-conservative. For this reason the natural unit of analysis would be the overall score in the cage, pooling the results of all animals in the cage. Hence, this reviewer would generally recommend single housing of animals. However, dosing by gavage would be expected to reduce the effects of these correlations relative to some other forms of dosing, e.g. dietary dosing. For consistency with other analyses the effects of these possible correlations are ignored, and the unit of analysis used in this report is the individual animal.

The Sponsor also indicates that “Blood samples for haematology investigations were collected from the first surviving 10 males and 10 females in each main study group in Weeks 53 and 78. Blood samples were taken prior to termination from all surviving main study animals during Week 103 or 104 or at necropsy for Group 4 males.” (page 15 of report) Since these were main study animals one might speculate if there was any effect on dose related tumorigenicity.

1.3.1.2. Survival Analysis:

Two main test statistics are provided, the log rank test and the so-called Wilcoxon test. The log rank tests puts equal weight on all events being assessed, while the Wilcoxon test weights them by the square of their rank in time, and thus places more weight on later events than does the log rank test. So the Wilcoxon test will generally be more sensitive to later separation of mortality than will be the log rank test. The number of such tests raises issues of multiple testing, but from the point of view of finding differences among treatment groups (i.e., reducing the probability of Type II error), this should be acceptable. Appendix 1 reviews the animal survival analyses in some detail. The Sponsor's analyses are summarized in Section 3.2.1.1. Appendix 2 presents an experimental Bayesian approach that allows nonproportional hazards.

1.3.1.3. Tests on Neoplasms:

The Sponsor provided the results of so-called Peto analyses. These require assessment of whether or not a tumor can be classified as fatal, a classification which can be difficult to accurately determine. Further, Peto analysis of observable tumors is also based on time of detection. Other tumors are defined as "incidental" and the incidence of such tumors are using tests stratified on time period. Until recently, most submissions to the CDER were analyzed with such Peto tests. However, the Society of Toxicological Pathology had a town hall meeting in June 2001 where this approach was criticized. The primary alternative discussed in the commentary on this meeting (STP Peto Working Group, 2002) is the poly-k modification of the Cochran-Armitage test of trend for tumor incidence. This is the method of analysis used in this report. Note that Appendix 4 provides an outline of an experimental Bayesian analysis of tumorigenicity.

1.3.1.4. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, which in turn necessitated an adjustment in experiment-wise Type I error. One, perhaps the usual, approach for two species, two gender, two year studies with testing for trend over four doses and comparing the high dose group to controls follows the Haseman-Lin-Rahman rules for the Peto analysis. Based on his extensive experience with such analyses, for pairwise tests between the high dose group and controls in two species, 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. For a standard chronic study in two species (i.e., mice and mice) study, based on simulations and their experience, 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. This is the adjustment used by the Sponsor. However, since this is a one species study, for rare tumors both the test of trend and the test of pairwise comparison between the high dose group and the appropriate control should be tested at a 0.05 (5%) level. The corresponding tests for common tumors should be tested at a 0.01 level. In this analysis we will use the observed incidence in the

vehicle control group to decide if a tumor is rare or common.

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 tumorigenicity when there actually is such a relation). Rahman and Lin (2008) showed these rules are also roughly applicable to the poly-k analyses used here. Note that strictly speaking, these rules only control the overall errors of the tests of trend over four doses for two genders, each with one marginal trend test and a corresponding test comparing the high dose to the control. It is not clear how the error rate would apply to other possible tests.

As discussed in Berger and Savage (1966) one can consider that the hierarchical nature of the Bayesian methods used here (as in Appendices 2 and 4), incorporate an automatic correction for multiplicity.

1.3.1.5. 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 (1994), quoting work by Haseman, have suggested that a survival rate of about 25 animals, out of 50 or more animals, 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. Note this criterion does seem to be satisfied.

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’ ” The values in the following two tables were taken from the Sponsor’s tables. Table 3 gives the mean change in body weight for each dose group and the final percent weight change relative to the weight change in the controls. Note that the criterion cited above seems to fail in male mice. Although this is a decision for the toxicologist, this may be evidence that the MTD was exceeded in the high dose group in males.

Table 3. Relative Group Mean Weight Change (g – compared to control)

Dose Label Nominal Dose mg/kg/day	Males		Females	
	Change	% from Control	Change	% from Control
Controls – 0	16.3		18.0	
Low - 100	15.3	-6.1%	22.5	25.0%
Medium - 250	14.7	-9.8%	17.2	-4.4%
High – 250/600	12.0	-26.4%	16.5	-8.3%

The Sponsor’s report summarizes the weight changes as follows: “Over the period of continuing growth (approximately to Week 52), females given 100 mg/kg/day gained more weight (non-statistically significant) than Controls. Bodyweight gains for males and higher dosage groups in females were similar to those of Controls to Week 52.

“Subsequently, bodyweight gain fluctuated in line with increasing and changing mortality and further apparent inter-group differences were considered not attributable to treatment with BIA 2-093.

“Over the duration of the study, males given 250 or 600 g/kg/day gained less weight than the Controls, but only achieving statistical significance for males given 600 mg/kg/day. Whereas, in the females, females given 100 mg/kg/day gained more weight than Controls and females given 250 or 600 mg/kg/day gained similar weight to Controls.” (page 37 of report)

The Sponsor’s report also states that: “Over the treatment period as a whole, the mean daily food intake was similar for animals given BIA 2-093 and Controls.

“During the first 16 weeks of the study females given 600 mg/kg/day had a lower mean daily food intake in comparison with Controls.

“From Week 53 males given 600 mg/kg/day had a statistically significantly higher daily food intake in comparison with Controls, this was also seen in males given 250 mg/kg/day and females given 600 mg/kg/day from Week 81 onwards.” (page 38 of report)

The following table, Table 4, displays the overall mean over weeks of the mean food consumption per treatment group. The values are taken from the Sponsor’s table of food consumption (page 38). This table is consistent with the Sponsor’s observation above.

Table 4. Means of Mean Food Consumption (g/animal/day)

Dose Label	Nominal Dose mg/kg/day	Males		Females	
		Mean	% from Control	Mean	% from Control
Control	0	6.4		5.1	
Low	100	6.3	-1.6%	5.2	1.9%
Medium	250	6.4	0%	5.1	0%
High	250/600	6.8	6.3%	5.2	1.9%

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. One way to assess this possibility is to measure mortality not associated with any identified tumor. Note this seems to be a new way to assess if the high dose is at the MTD. Table 5 below indicates the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors:

Table 5. Natural Death or Accident with No Identified Tumor

Group Label	Dose Mg/kg/day	Males		Females	
		Died w/o tumor	Other	Died w/o tumor	Other
Control	0	3	51	7	47
Low	100	5	49	6	48
Medium	250	0	54	5	49
High	250/600	0	54	1	53

To compare the incidence of deaths without tumors we can specify the usual survival tests where animals that die with a tumor or are sacrificed are considered as censored. The remaining animals are those that die prior to developing a tumor. If the MTD is exceeded we would expect a dose related excess toxicity, resulting in a dose related trend in these deaths. In both genders there is no evidence of such an increasing trend over dose in non-tumor related deaths. If one wishes to support this by a statistical analysis, one appropriate analysis would be to use the usual time to event tests to investigate if there are dose related differences between treatment groups (Male: logrank $p=0.0254$, Wilcoxon $p=.0130$, Female: logrank $p=0.6373$, Wilcoxon $p=0.7545$). In both genders there is actually a decreasing tendency for non-tumor related deaths over levels of dose. In male mice this does appear to be due to competing risks. Although this evidence is hardly conclusive, and this is a decision for the toxicologist, this may be evidence that in both genders the MTD was not exceeded, and possibly was not achieved.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

According to the Sponsor, the objective of this study was to assess the influence of Eslicarbazepine acetate on tumour formation when administered by daily gavage to Crl:CD-1 (ICR) BR VAF/PLUS derived strain mice for up to 103 weeks in males and 104 weeks in females. There were four different treatment groups: a vehicle control (0.5% hydroxypropyl-methyl cellulose) and three doses of Eslicarbazepine acetate at 100, 250, or 600 mg/kg/day.

2.2. Data Sources

One SAS transport file was provided by the Sponsor and placed in the CDER electronic data room (edr):

093-830.xpt containing the single SAS data set fda2a.sas7bdat.

Note that, at the request of the toxicologist, a number of the tumor types listed in this data were combined for analysis. They are denoted by lower case names in tumor name field in the tables of neoplasms in this report.

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. Study FLG0024: 104 Week Oral (Gavage) Carcinogenicity Study in the Mouse

STUDY DURATION: 103/104 Weeks.

DOSING START DATE: September 29, 2003

TERMINAL SACRIFICE (NECROPSY) DATES: Males: 19 & 20 September, 2005.

Females: 30 September-3 October, 2005

STUDY ENDING DATE (Completion of experimental work): July 4, 2006.

MOUSE STRAIN: Crl:CD-1 (ICR) BR VAF/PLUS derived strain mice .

ROUTE: Gavage

For this study 54 Crl:CD-1 (ICR) BR VAF/PLUS derived strain mice of each gender were allocated to four different treatment groups: a vehicle control (0.5% hydroxypropylmethylcellulose) and doses of Eslicarbazepine acetate at 100, 250, or 600 mg/kg/day. Note that the nominal high dose group, the 600 mg/kg/day group was actually dosed at 250 mg/kg/day for the first week of the study “in an attempt to alleviate the severity of the clinical signs” (page 14 of report). In addition to the 54 animals within each gender within each of the dose groups 1-4, corresponding satellite groups 5-8 were defined for toxicokinetic analysis. Group 5 controls were assigned 6 animals per gender, while groups 6-8, i.e., Eslicarbazepine acetate groups at 100, 250, or 600 mg/kg/day were assigned 12 animals per group.

The Sponsor also indicates that “Blood samples for haematology investigations were collected from the first surviving 10 males and 10 females in each main study group in Weeks 53 and 78. Blood samples were taken prior to termination from all surviving main study animals during Week 103 or 104 or at necropsy for Group 4 males.” (page 15 of report) Since these were main study animals one might speculate if there was any effect on dose related tumorigenicity.

Animals were dosed once daily by oral gavage, 7 days a week for at least 103 weeks for males and 104 weeks for females until the day before necropsy.

According to the Sponsor: “Males were housed in groups of 3 and females were housed in groups of 5, 4 and 3 in grid bottomed cages suspended over paper-lined trays. Due to the severity of the clinical signs being recorded for animals given 600 mg/kg/day, a coarse shredded nesting material was added to the cages of these animals from Day 54 onwards.” (page 24 of report) As discussed in Section 1.3.1.1 above, it is possible that such multiple housing may affect the results of the analysis.

“A pelleted diet, (b) (4) and mains tap water in polycarbonate or polypropylene bottles were freely available.” (page 24 of report)

The Sponsor further noted that “The route of administration corresponds to a proposed human therapeutic route.” (page 18 of report) “The dose levels of 100, 250 and 600 mg/kg/day were selected on the basis of a 13-week preliminary study conducted at (b) (4) (b) (4) Study No. NML 1097 (1)), which used BIA 2-093 dose levels of 150, 350, 500 and 650 mg/kg/day.

“A high dose level of 600 mg/kg/day was selected for this mouse carcinogenicity study because the very marked increases in liver weight observed at 650 mg/kg/day in the preliminary study (1); with a mean value of 150 % to 160 % of Control liver weight for animals given 650 mg/kg/day were considered unsuitable for extended administration periods (2). Dose escalation from 250 mg/kg/day to 600 mg/kg/day was used to try and alleviate the severity of the clinical signs (prostration, subdued behaviour, unsteady gait *etc*) observed in the first week of dosing of

the 13-week study (which were associated with the premature death of animals at 650 mg/kg/day).

“Based on single dose data, exposure (AUC_{0-24h}) of the primary human metabolite (BIA 2-005) on Day 1 of the study at the proposed high dose level was expected to be around 3 to 5 times the anticipated human therapeutic level for male mice and around 2.5 to 4 times for female mice.

“The intermediate dose level of 250 mg/kg/day was expected to produce similar, but slightly less severe, early clinical signs to the high dose level but to produce only moderate effects upon liver weight / hypertrophy. The low dose of 100 mg/kg/day was expected to be a No Observable Effect Level (NOEL).

“Dose administration began on 29 September 2003 and the necropsies were performed on 19 and 20 September 2005 for males and 30 September 2005 to 3 October 2005 for females.” (page 18-19 of report)

3.2.1.1. Sponsor’s Results and Conclusions

This section will present a summary of the Sponsor’s analysis on survivability and tumorigenicity in mice.

Survival analysis:

According to the Sponsor: “Due to low survival amongst Group 1 males, the males were terminated slightly earlier than planned during week 104. As shown [below] ... 150 males and 140 females died before the terminal kill. In the males there was some evidence of higher survival in Group 2 (p<0.05) but none of the other treated groups groups showed any difference from control and there was no significant dose related effect.”

“In females all groups had similar survival rates and so there were no significant differences.” (page 1680 of report)

Table 6: Sponsor’s Table 2: Percentage survival at select weeks.

Group Sex	Dose level (mg/kg/day)	Percentage survival at selected weeks				
		12	28	52	80	103/104#
1M	0	100	100	94	59	20
2M	100	98	96	93	74	44*
3M	250	98	96	93	63	33
4M	250/600	94	91	72	50	24
1F	0	100	98	89	59	39
2F	100	100	100	93	74	39
3F	250	100	98	96	69	35
4F	250/600	94	93	83	76	31

- statistically analysed *=p<0.05

“There was no overall effect on survival trends in females. However, for males given 100 mg/kg/day there were a higher number of animals surviving until the end of the study in comparison with Controls, however, this was considered to be fortuitous and not test article related.” (page 35 of report)

“Approximately 60 % to 70 % of decedent males and 70 % to 90 % of decedent females were euthanased due to their clinical condition; the remainder being found dead without previous indication of a decline in clinical condition.” (page 35 of report)

Tumorigenicity analysis:

The Sponsor cites the following as the most common tumors:

“Harderian glands, benign adenoma – 43 animals, 4 fatal
Liver, benign adenoma – 94 animals, 4 fatal
Liver, malignant hepatocellular carcinoma – 93 animals, 45 fatal
Lungs, malignant adenocarcinoma – 29 animals, 17 fatal
Lungs, benign adenoma – 63 animals, 1 fatal
Haemopoietic, malignant follicular cell lymphoma – 86 animals, 60 fatal
Haemopoietic, malignant histiocytic sarcoma – 18 animals, 12 fatal” (page 1680 of report)

Further, the Sponsor cites the main positive findings as below:

- “Liver, benign adenoma: There were significant increasing trends for both sexes ($p < 0.01$ in males and $p < 0.001$ in females). In males, Groups 3 and 4 showed significant increases from the control ($p < 0.01$), whereas in females, the significant increase was restricted to Group 4 ($p < 0.001$).
- “Liver, malignant hepatocellular carcinoma: There were significant increasing trends for both sexes ($p < 0.001$). In males, Groups 3 and 4 showed significant increases from the control ($p < 0.001$), whereas in females, the significant increase was restricted to Group 4 ($p < 0.001$).
- “Any site, malignant tumors: There was a significant increasing trend for the males ($p < 0.001$), but not for the females. When combined with the benign tumors, the male trend was slightly less convincing ($p < 0.01$), whereas the females were still non significant. However, the effects seen here seem entirely due to liver tumours since when that organ is removed from all site analysis there are no significant effects.
- “Multiple sites, malignant tumors: There was a significant increasing trend in the females ($p < 0.001$) but this was not seen in the males. However the trend for males was significant ($p < 0.001$) when the benign and malignant tumors were considered together.

Again, the effects seen here seem entirely due to liver tumours since when that organ is removed from the multiple site analysis there are no significant effects.

“There was also a negative finding in the lungs as follows.

- “Lungs, benign adenomas and malignant adenocarcinoma: There were fewer tumours in the high dose group than in control for both sexes. Although the trend effect was not statistically significant in the males, it was in the females ($p < 0.01$).” (page 1681 of 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 mice.

Survival analysis:

The following tables (Table 7 for male mice, Table 8 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 that survived to the end of the interval.

Table 7. Summary of Male Mice Survival (Eslicarbazepine: daily dose)

Period (Weeks)	Vehicle Control	Low 100 mg/kg	Medium 500 mg/kg	High 2000 mg/kg
0-52	3/54 94.4%	4/54 92.6%	4/54 92.6%	15/54 72.8%
53-78	16/51 64.8%	10/50 74.1%	13/50 68.5%	6/39 61.1%
79-91	12/35 42.6%	6/40 63.0%	11/37 48.1%	14/33 35.2%
92-106	12/23 20.4%	10/34 44.4%	8/26 33.3%	6/19 24.1%
Terminal 106	11	24	18	13

¹ number deaths / number at risk

² per cent survival to end of period.

Table 8. Summary of Female Mice Survival (Eslicarbazepine: daily dose)

Period (Weeks)	Vehicle Control	Low 100 mg/kg	Medium 500 mg/kg	High 2000 mg/kg
0-52	5/54 90.7%	3/54 94.4%	2/54 96.3%	9/54 83.3%
53-78	14/49 64.8%	10/51 75.9%	11/52 75.9%	4/45 75.9%
79-91	8/35 50.0%	9/41 59.3%	11/41 55.6%	13/41 51.8%
92-107	6/27 38.9%	13/32 35.2%	11/30 35.2%	11/28 31.5%
Terminal 107-108	21	19	19	17

¹ number deaths / number at risk

² per cent survival to end of period.

The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table 9 below. As noted in Section 1.3.1.2 above Wilcoxon test places more weight on later events than does the log rank test. So the Wilcoxon test will be more sensitive to later separation of mortality than will be the log rank test.

Table 9. Statistical Significances of Tests of Homogeneity and Trend in Survival

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Homogeneity over Groups 1-4	0.0215	0.0189	0.5682	0.3259
Trend over Groups 1-4	0.1630	0.0572	0.5499	0.6809
Comparison of High and Low	0.7171	0.7474	0.8481	0.6389

Kaplan-Meier estimated survival curves across dose groups for each gender in each study are given in Figures A.1.1-A.1.2 of Appendix 1. These curves help in understanding the results of the tests above. In female mice, the Kaplan-Meier curves are generally intertwined, with no strong evidence of trend or lack of homogeneity over the dose groups (all $p \geq 0.3259$). In male mice survival in the high dose group is lower than the other groups, particularly early in the study, resulting in the statistically significant tests of the null hypothesis of homogeneity (Logrank $p=0.0215$, Wilcoxon $p=0.0189$). However this separation in survival is generally earlier in the study. By the end of the study the vehicle control and the Eslicarbazepine high dose group are fairly close, which explains the statistically non-significant result comparing the high dose group to the control (both $p \geq 0.7171$). The wandering of the survival curves in male mice seems to be the reason for the somewhat ambiguous results of the simple test of trend (Logrank $p=0.1630$, Wilcoxon $p=0.0572$).

One way to at least roughly assess departure from trend in survival among dose groups would be to use a semi-parametric, proportional hazards model. One could then fit the doses as

levels of a factor and as a continuous covariate and compute the differences of the corresponding log partial likelihoods. Under the proportional hazards assumptions, these will be asymptotically chi-square with 3 degrees of freedom. These tests were barely statistically significant in males ($p = 0.0457$) and highly non-significant in females ($p = 0.9939$). The assumptions of the proportional hazard model are not met, but these tests may still provide some guideline in that there is equivocal evidence of a departure from simple trend in male mice.

Appendix 2 includes an experimental Bayesian analysis that allows nonproportional hazards, by defining a period wise constant hazard function for each dose. It indicates that the high dose group has greater hazard than the remaining doses during the first year of the study, particularly in male mice.

Tumorigenicity analysis:

The four dose groups are the vehicle controls and the three actual treatment groups with nominal dose levels of Eslicarbazepine acetate at 100, 250, and 250/600 mg/kg/day by daily gavage. The latter three dose groups, Sponsor groups 2-4, are also labeled as Low, Medium, and High, respectively. Table 10 below lists all tumors that are potentially statistically significant (any $p \leq 0.05$). Appendix 3 includes complete incidence tables and the results of corresponding poly-k tests of trend and pairwise differences with the vehicle control. In this study these were all liver tumors. As noted in Section 1.3.1.4., this is a one species two gender study, so to preserve overall Type I error to a rough 10%, for rare tumors (with incidence $\leq 1\%$) both the test of trend and the test of pairwise comparison between the high dose group and the appropriate control should be tested at a 0.05 (5%) level. The corresponding tests for common tumors (with incidence $> 1\%$) should be tested at a 0.01 level.

In Table 10, note that the first p-value provides the results of the overall poly-k test of trend. The poly-k test modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The last three columns present the results of tests between the control group and each of the high dose group, the medium dose group, the low dose group respectively.

Using the incidence in the control group to define whether the tumor is classified as rare or common, all these tumors in male mice would be classified as common and all tumors in female mice would be classified as rare. In male mice the test of trend over dose was statistically significant ($p=0.0086 < 0.01$), as was the test in hepatocellular carcinoma and pooled adenoma and carcinoma ($p < 0.00005 < 0.01$). In female mice, results were even more highly statistically significant, for all three tests of trend in adenoma, carcinoma, and pooled adenoma and carcinoma (all $p < 0.00005 < 0.05$). In male mice the test comparing the high dose group to the vehicle control group in adenomas was not quite statistically significant ($p = 0.0171 > 0.01$). However, in both genders the tests comparing the high dose group to controls were highly statistically significant in both carcinomas and pooled adenomas and carcinomas ($p = 0.00005 < 0.01, 0.05$). The test comparing the high dose group to controls in adenomas in female mice was also highly statistically significant ($p = 0.0001 < 0.05$). In male rats the pairwise test of the

medium dose group compared to the vehicle controls in adenoma was also statistically significant ($p = 0.0078 < 0.01$), as was the pairwise test in carcinoma and pooled adenoma and carcinoma (both $p < 0.00005 < 0.01$). No other pairwise comparison was even statistically significant at a nominal 0.05 level.

Table 10. Potentially Statistically Significant Neoplasms in Mice

Tumor	Veh	Low	Med	High	Trend	Veh vs Low	Veh vs Med	Veh vs High
Male Mice								
Liver								
ADENOMA - BENIGN	12	18	26	23	0.0086	0.2653	0.0078	0.0171
HEPATOCELLULAR CARCINOMA	5	10	29	32	0.0000	0.2848	0.0000	0.0000
Adenoma/Hepato. Carc.	15	24	38	40	0.0000	0.1784	0.0000	0.0000
Female Mice								
Liver								
ADENOMA - BENIGN	0	2	1	12	0.0000	0.2680	0.5143	0.0001
HEPATOCELLULAR CARCINOMA	0	0	2	15	0.0000	1.0000	0.2609	0.0000
Adenoma/Hepato. Carc.	0	2	2	24	0.0000	0.2680	0.2609	0.0000

Appendix 4 includes a Bayesian analysis of tumorigenicity. This analysis includes assessments of the probability that a parameter is 0 or not. Using this analysis, similar to the results above, in the livers of males there is strong evidence of a dose related trend in hepatocellular carcinoma and pooled adenoma and hepatocellular carcinoma in the liver, as well as an assessment of the probability that the effect of the high dose is the same as the control (i.e., for all these $p \leq 0.00005$). In females, as with males there is strong evidence of a dose related trend in hepatocellular carcinoma and pooled adenoma and hepatocellular carcinoma in the liver (i.e., for $p = \text{probability of trend}$, both $p \leq 0.00005$). The same comment holds for the difference between the high dose group and controls for hepatocellular carcinoma in the liver. For pooled adenoma and hepatocellular carcinoma in the liver the probability of a difference between the high dose group and controls is 0.002.

It should be emphasized that the probabilities in the Bayesian analysis cited above are not significance levels. In the poly-k analysis in Table 10 and Appendix 3, significance levels are defined as the estimated probability of a result as “extreme” or more extreme than that which was observed, assuming there is no effect of dose. In the Bayesian approach, if we assume the prior distributions represent reasonable expressions of uncertainty about the parameters, then the probabilities cited above represent reasonable expressions of the probability that the parameter is actually zero.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

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 above.

APPENDICES:**Appendix 1. Survival Analysis**

Simple summary life tables in mortality are presented in the report (Tables 8 and 9), above. Kaplan-Meier estimated survival curves across dose groups for each gender in each study are displayed in Figures A.1.1-A.1.2 below. These plots include 95% confidence intervals around each curve (colored area around each curve). The plots are also supported by tests of homogeneity over the four dose groups, simple tests of trend in survival, and finally a pairwise comparison between the high dose and the control groups. The statistical significances of the tests of differences in survival across treatment groups using the log rank and the so-called Wilcoxon test are given in Table A.1.1. below. One might note that the log rank tests puts equal weight on all events, while the Wilcoxon test weights them by the square of the time rank, and thus places more weight on later events than does the log rank test. So the Wilcoxon test will be more sensitive to later separation of mortality than will be the log rank test.

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

	Males		Females	
	Log Rank	Wilcoxon	Log Rank	Wilcoxon
Homogeneity over Groups 1-4	0.0215	0.0189	0.5682	0.3259
Trend over Groups 1-4	0.1630	0.0572	0.5499	0.6809
Comparison of High and Low	0.7171	0.7474	0.8481	0.6389

In female mice, the Kaplan-Meier curves are generally intertwined, with no strong evidence of trend or lack of homogeneity over all dose groups (all $p \geq 0.3259$). In male mice survival in the high dose group is lower than the other groups, particularly early in the study, resulting in the statistically significant tests of the null hypothesis of homogeneity (Logrank $p=0.0215$, Wilcoxon $p=0.0189$). However this separation in survival is generally earlier in the study. By the end of the study the vehicle control and the Eslicarbazepine high dose group are fairly close, which explains the statistically non-significant result comparing the high dose group to the control (both $p \geq 0.7171$). The wandering of the survival curves in male mice seems to be the reason for the somewhat ambiguous results of the simple test of trend (Logrank $p=0.1630$, Wilcoxon $p=0.0572$).

One way to at least roughly assess departure from trend would be to use a semi-parametric, proportional hazards model. One could then fit the doses as levels of a factor and as a continuous covariate and compute the differences of the corresponding log partial likelihoods. Under the proportional hazards assumptions, these will be asymptotically chi-square with 3 degrees of freedom. These tests were barely statistically significant in males ($p = 0.0457$) and highly non-significant in females ($p = 0.9939$). The assumptions of the proportional hazard model are not met, but these tests may still provide some guideline in that there is equivocal

evidence of a departure from simple trend in male mice. Visually this is clear in figure A.1.1 below.

Figures A.1.1 and A.1.2, below, display these Kaplan-Meier estimated survival curves for the two genders.

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

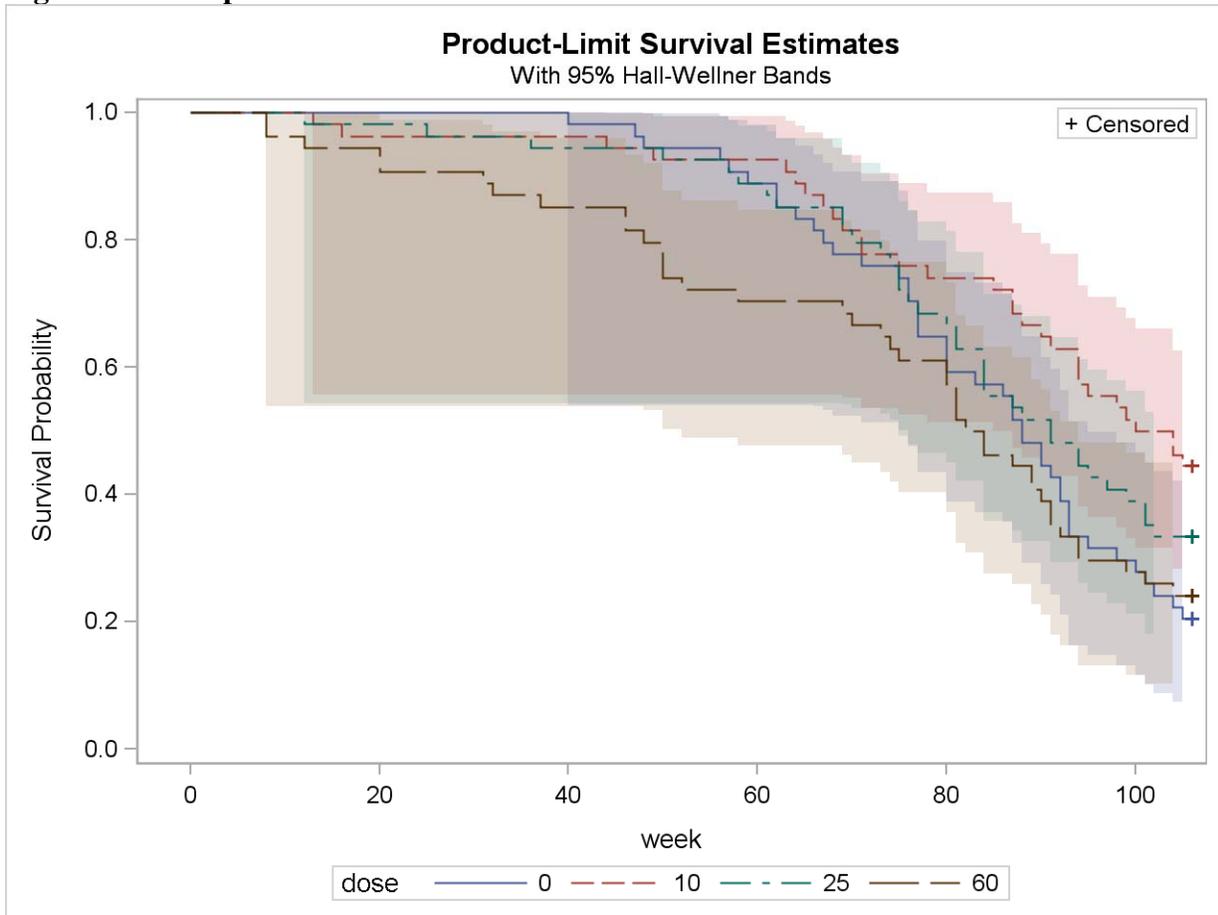
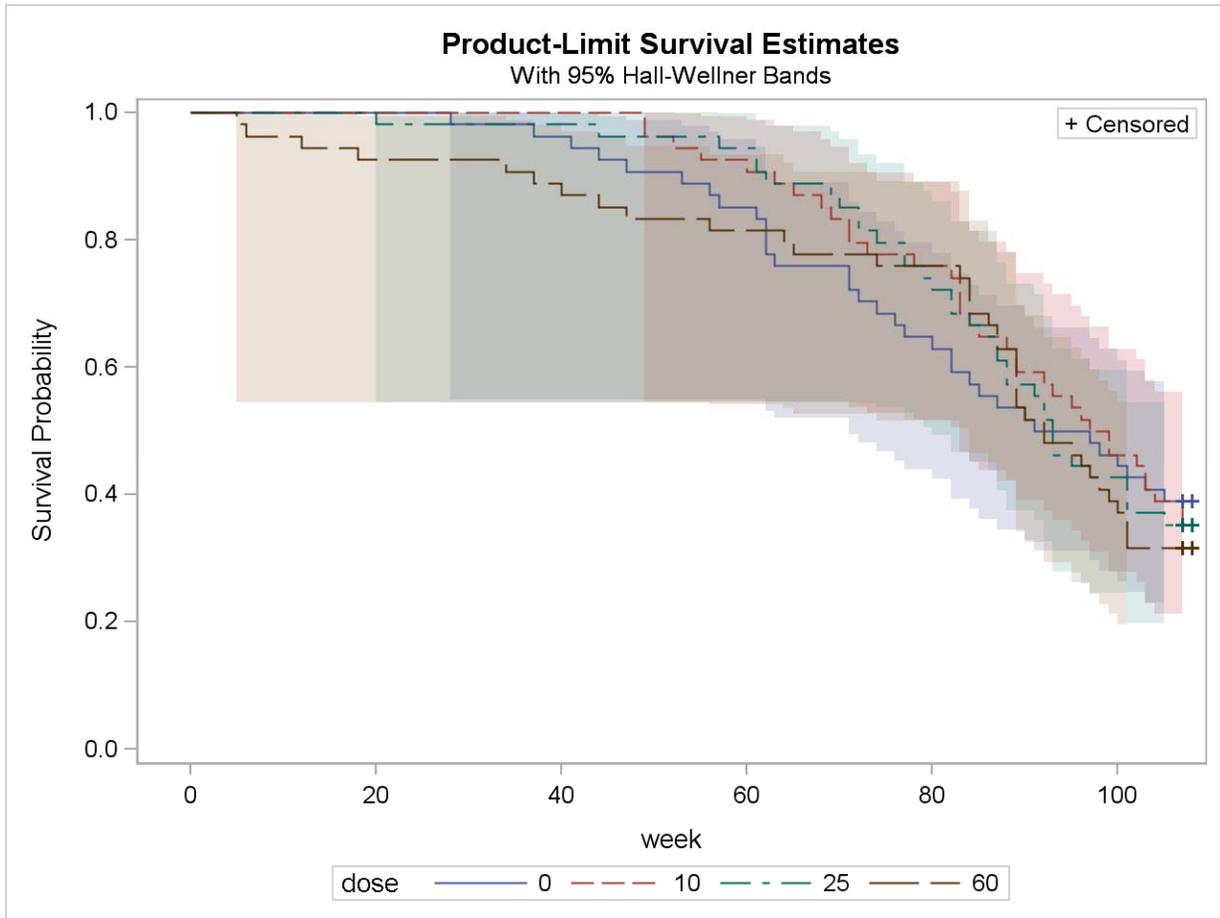


Figure A.1.2 Kaplan-Meier Survival Curves for Female Mice



Appendix 2. Bayesian Analysis of Survival

Let $S(t)$ be the survival function, i.e., with T denoting the random survival time,

$$S(t) = Pr(T > t),$$

and $f(t)$ the density of T . The instantaneous hazard function is $h(t) = f(t)/S(t)$ with cumulative hazard up to time t :

$$H(t) = \int_0^t h(u) du$$

So $f(t) = h(t) S(t)$. Also $\log(S(t)) = -H(t)$, so $S(t) = e^{-H(t)}$. Then $f(t) = h(t) e^{-H(t)}$.

The standard Cox regression form of the proportional hazards model for survival specifies the hazard function:

$$h(t | x) = h_0(t) \exp(x'\beta),$$

where $h_0(t)$ is the same across treatment groups. Then dose group only enters $h(t | x)$ through $\exp(x'\beta)$, and for each treatment group d , $H(t)$ is proportional to $\exp(x'\beta_d)$. Since $\log(S(t)) = -H(t)$, the corresponding survival curves will tend to track each other without intersecting.

It is clear from the plots in Section 1 that in this case the observed survival curves do intersect. A typical frequentist approach is to introduce time dependent covariates to adjust for such intersections. Arguably a more sensible approach is to allow the estimated baseline hazard $h_0(t)$ to differ across treatment groups. Perhaps the simplest Bayesian model would postulate a within interval constant baseline hazard, but a different hazard for each treatment group. That is, suppose the time axis can be partitioned as $(a_1=0, a_2]$, $(a_2, a_3]$, . . . , $(a_T, a_{T+1}]$. Assume a constant baseline hazard λ_{dj} for observations from treatment group d , out of a total of g groups, to indicate the baseline hazard in interval $(a_j, a_{j+1}]$.

Let t_i = time to failure or censoring for the i th subject and suppose it is in the interval $(a_{j-1}, a_j]$. So the integrated cumulative baseline hazard for this subject can be written as:

$$H_{od}(t_i) = \int_0^{t_i} h_{d0}(u) du = \left\{ \sum_{k=1}^{j-1} \lambda_{dk} (a_k - a_{k-1}) + \lambda_{dj} (t_i - a_{j-1}) \right\},$$

with instantaneous hazard $h_{od}(t_i) = \lambda_{dj}$. Note that the cumulative hazard will be represented as piece-wise linear function, with the cumulative hazard increasing at the constant rate λ_{dj} within each interval $(a_{k-1}, a_k]$.

The likelihood for subject i in dose group d can then be written as:

$$L_i(\lambda, \beta) \propto \begin{cases} e^{-H_{od}(t_i)} & \text{if } i\text{th subject is censored at time } t_i \\ \lambda_{dj} e^{-H_{od}(t_i)} & \text{if } i\text{th subject fails at time } t_i \end{cases}$$

Because this looks like a sample of exponential interarrival times we would expect the simple fail/not fail distributions to correspond to Poisson random variables.

For subject i censored or failed at time t_i , let $\gamma_{idk} = \begin{cases} \lambda_{dk}(a_k - a_{k-1}) & \text{for } t_i > a_k \\ \lambda_{dj}(t_i - a_{j-1}) & \text{for } a_{j-1} \leq t_i < a_j \\ 0 & \text{otherwise} \end{cases}$

Note that for intervals above a_j , the term $\gamma_{idk} = 0$, so for those intervals $\exp(-\gamma_{idk})$ does not contribute to the product. Then $S_d(t) = e^{-H(t)} = \prod_{k=1}^T \exp(-\gamma_{idk})$. Thus, for subject i in group d , the likelihood can also be written as:

$$L_i(\lambda, \beta) \propto \begin{cases} \prod_{k=1}^T \exp(-\gamma_{idk}) & \text{if } i\text{th subject is censored at time } t_i \\ \gamma_{idj} \prod_{k=1}^T \exp(-\gamma_{idk}) & \text{if } i\text{th subject fails at time } t_i \end{cases}$$

Note this corresponds to the likelihood of T independent Poisson random variables with mean γ_{dik} where all responses are zero except at time j with the occurrence of a failure in the j th interval $(a_{j-1}, a_j]$. Since all responses are 0 or 1, this is only a computational convenience but allows estimation of the appropriate parameters. As an aside, note that it would be easy to incorporate other individual level covariates besides dose in the same manner as a typical semi-parametric cox regression, i.e. for the i th subject with covariates x_i , merely by replace each λ_{dj} by $\lambda_{dj} \exp(\beta^t x_i)$. However that is not done in this analysis.

The time intervals used for the baseline hazards in this analysis match those used in Tables 7 and 8. These are long intervals, but for robustness of results we need to have a sufficient number of observations to estimate the within dose, within time period, probability of mortality. Sacrifice or accidental death is treated as a reduction in the risk set, but not as a mortality event. A gamma prior on the within treatment group, within time period, hazard would be skewed to the right and would seem to be an appropriate choice of family for the baseline hazards λ_{dj} .

To reflect the expectation of an increasing hazard, for period j , $j=1, \dots, 4$, we specify a gamma prior on the λ_{dj} with location 0.25 and scale parameter $25*j$. This implies a prior scaled hazard with mean $100*j$ and variance $400*j$. This would seem to be a relatively noninformative prior.

Table A.2.1, below, summarizes the estimated posterior distributions of the treatment parameters. Time in study was measured in weeks. Since the time intervals are of different lengths, and to increase readability in the tables the mean in the tables below correspond to $100*\lambda_{dj}$, i.e., 100 times the hazard. This is the Poisson mean times 100 divided by the length of the interval, and might be called “normalized hazard means”. The standard deviation of the

values and the 0.25%, median (i.e. 50%), and 97.5% quantiles are also presented. One measure of difference between doses is, within each time period, to compare this mean to the average of the other three dose groups. For assessing overall treatment differences this is the most interesting measure. The mean difference, its standard deviation, and the corresponding quantiles are also presented. Note that the actual probability that the differences are in the range between the 2.5% and 97.5% percentiles is 0.95, defining a so-called 95% credible interval. Thus, if 0 is not in that interval, we know that the probability that each parameter is “close” to the mean of the remaining parameters is less than 0.05, providing relatively strong evidence that they are different.

Table A.2.1 Posterior Summaries of Treatment Parameters in Males

Weeks	Dose	Normalized mean		Quantiles			Difference			Quantiles		
		Mean	std dev	2.5%	50%	97.5%	mean	sd	diff	2.5%	50%	97.5%
0-52	0	0.113	0.063	0.03	0.10	0.26	-0.179	0.087	-0.34	-0.18	0.00	
	100	0.153	0.074	0.04	0.14	0.33	-0.127	0.094	-0.30	-0.13	0.08	
	250	0.152	0.073	0.04	0.14	0.32	-0.128	0.094	-0.30	-0.13	0.07	
	600	0.573	0.148	0.32	0.56	0.90	0.433	0.153	0.17	0.42	0.77	
53-78	0	1.341	0.331	0.78	1.31	2.05	0.502	0.365	-0.16	0.48	1.28	
	100	0.818	0.254	0.39	0.79	1.38	-0.196	0.305	-0.76	-0.21	0.45	
	250	1.082	0.296	0.59	1.05	1.74	0.157	0.338	-0.45	0.14	0.87	
	600	0.617	0.250	0.23	0.58	1.19	-0.463	0.303	-1.00	-0.48	0.19	
79-91	0	2.708	0.742	1.46	2.64	4.35	0.461	0.840	-1.04	0.41	2.25	
	100	1.092	0.433	0.42	1.03	2.09	-1.694	0.616	-2.86	-1.71	-0.44	
	250	2.268	0.671	1.14	2.21	3.75	-0.125	0.784	-1.55	-0.16	1.51	
	600	3.380	0.868	1.89	3.31	5.24	1.358	0.941	-0.32	1.29	3.36	
92-term- inal	0	3.480	1.039	1.77	3.38	5.80	1.574	1.112	-0.35	1.49	4.00	
	100	2.015	0.628	0.98	1.95	3.41	-0.379	0.796	-1.87	-0.41	1.27	
	250	2.047	0.716	0.89	1.96	3.68	-0.337	0.856	-1.90	-0.39	1.46	
	600	1.657	0.722	0.56	1.55	3.34	-0.857	0.862	-2.38	-0.92	1.02	

In males only the 95% credible intervals for the difference of the high dose group (600 mg/kg/day) normalized mean from the average of the three other dose group normalized means in the first period, weeks 0-52, (i.e., 0.17 to 0.77) and the difference of the low dose group in the third period, weeks 79-91, from the three others (i.e., -2.85 to -0.44) do not include 0. Note the first credible interval corresponds to an increase in hazard over the average hazard of the others groups while the second to a decrease in hazard.

The following table summarizes the distribution of the difference of each of the 100, 250, and 600 mg/kg/day dose normalized means with the 0 dose control normalized mean. Conclusions are virtually identical to those above.

Table A.2.2 Comparison of Treatment Parameters to Control in Males

Weeks	Dose	Normalized mean			Quantiles		
		Difference	std dev	2.5%	50%	97.5%	
0-52	100	0.039	0.097	-0.15	0.04	0.24	
	250	0.039	0.097	-0.15	0.04	0.24	
	600	0.459	0.160	0.17	0.45	0.80	
53-78	100	-0.524	0.416	-1.37	-0.51	0.27	
	250	-0.259	0.445	-1.14	-0.26	0.61	
	600	-0.724	0.416	-1.56	-0.72	0.08	
79-91	100	-1.616	0.857	-3.41	-1.58	-0.04	
	250	-0.440	0.998	-2.45	-0.43	1.51	
	600	0.672	1.145	-1.55	0.67	2.95	
91-term-inal	100	-1.465	1.216	-4.03	-1.40	0.76	
	250	-1.433	1.261	-4.08	-1.38	0.92	
	600	-1.823	1.265	-4.46	-1.78	0.59	

Table A.2.3, below, summarizes the estimated posterior distributions of the normalized treatment parameters, similar to Table A.2.1 above.

Table A.2.3 Posterior Summaries of Treatment Parameters in Females

Weeks	Dose	Normalized mean			Quantiles			Difference		Quantiles		
		Mean	std dev	2.5%	50%	97.5%	mean	sd diff	2.5%	50%	97.5%	
0-52	0	0.186	0.081	0.06	0.17	0.37	0.006	0.094	-0.16	-0.00	0.21	
	100	0.111	0.062	0.03	0.10	0.26	-0.093	0.079	-0.24	-0.10	0.08	
	250	0.079	0.053	0.01	0.07	0.21	-0.136	0.074	-0.27	-0.14	0.02	
	600	0.349	0.114	0.16	0.34	0.61	0.223	0.120	0.02	0.21	0.49	
53-78	0	1.245	0.329	0.69	1.22	1.97	0.562	0.356	-0.07	0.54	1.33	
	100	0.812	0.252	0.40	0.79	1.37	-0.015	0.295	-0.55	-0.03	0.61	
	250	0.873	0.261	0.45	0.84	1.45	0.065	0.302	-0.48	0.05	0.71	
	600	0.365	0.178	0.10	0.34	0.79	-0.612	0.241	-1.06	-0.62	-0.11	
79-91	0	1.674	0.580	0.74	1.61	2.98	-0.421	0.681	-1.64	-0.46	1.04	
	100	1.691	0.558	0.78	1.63	2.95	-0.398	0.666	-1.59	-0.43	1.01	
	250	2.053	0.613	1.03	1.99	3.42	0.084	0.707	-1.19	0.04	1.59	
	600	2.541	0.676	1.40	2.48	4.02	0.735	0.756	-0.62	0.69	2.34	
92-term-inal	0	1.335	0.531	0.51	1.27	2.56	-1.094	0.675	-2.33	-1.13	0.33	
	100	2.300	0.685	1.15	2.23	3.82	0.192	0.796	-1.25	0.16	1.89	
	250	2.550	0.757	1.30	2.47	4.24	0.525	0.851	-0.98	0.47	2.35	
	600	2.438	0.757	1.20	2.36	4.13	0.377	0.851	-1.13	0.33	2.20	

In females only the 95% credible intervals for the difference of the high dose group (600 mg/kg/day) from the average of the three others in the first period (i.e., 0.02 to 0.49) and the difference of the high dose group from the same average in the second period others (i.e., -1.06 to -0.11) do not include 0. Again, the first credible interval corresponds to an increase in hazard over the average hazard of the others groups while the second to a decrease in average hazard.

Table A.2.4 Comparison of Treatment Parameters to Control in Females

Weeks	Dose	Normalized mean		Quantiles		
		Difference	std dev	2.5%	50%	97.5%
0-52	100	-0.074	0.102	-0.29	-0.07	0.12
	250	-0.107	0.097	-0.31	-0.10	0.07
	600	0.163	0.140	-0.10	0.16	0.46
53-78	100	-0.433	0.416	-1.28	-0.42	0.35
	250	-0.373	0.420	-1.22	-0.37	0.44
	600	-0.880	0.374	-1.67	-0.86	-0.19
79-91	100	0.017	0.807	-1.59	0.02	1.62
	250	0.379	0.841	-1.29	0.38	2.04
	600	0.867	0.891	-0.88	0.86	2.66
92-term- inal	100	0.964	0.870	-0.71	0.95	2.74
	250	1.215	0.921	-0.52	1.19	3.11
	600	1.103	0.920	-0.63	1.08	3.00

In females the only 95% credible interval not containing 0 is the comparison between the high dose group and controls in the second period (-1.67 to -0.19), corresponding to decrease in hazard compared to the hazard of the control group.

Note this is an experimental approach. The appropriateness of this analysis and its generalizations is a topic for further research. Another alternative, but very experimental, Bayesian approach to hypothesis testing is illustrated in Appendix 4 below.

These analyses were implemented in WinBUGS 1.4.3 (see Lunn et al, 2000)

Appendix 3. Poly-k Tumorigenicity Analysis

Tables A.3.1-A.3.3 below, display the tumor incidence over the control and the three actual dosing groups, as well as the p-values of the poly-k (here with k=3) tests of trend in dose and pairwise comparisons to the vehicle controls. The first p-value provides the results of the overall poly-k test of trend. The poly-k test modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The last three columns present the results of tests between the control group and each of the high dose group, the medium dose group, the low dose group respectively.

As noted in the report, at the Society of Toxicological Pathology “town hall” meeting in June 2001 the poly-k modification of the Cochran-Armitage test of trend seemed to have been recommended over the so-called Peto tests. The tests used here are small sample exact tests, which assume all marginal totals are fixed, a debatable assumption. To adjust for the multiplicity of tests, tentatively, the Haseman-Lin-Rahman rules discussed in Section 1.3.1.4. of the report seem to apply, here with the modification for a single species study.

Table A.3.1 below lists tumors that have any p-value less than 0.05, in this case, all liver tumors. Applying the Haseman-Lin-Rahman rules for a single species study, rare tumors (incidence $\leq 1\%$) could be considered statistically significant if the observed p-value is 0.05 or less, while common tumors (incidence $> 1\%$) would be considered as statistically significant if the observed p-value is 0.01 or less. Using the incidence in the control group to define whether the tumor is classified as rare or common, all these tumors in male mice would be classified as common and all tumors in female mice would be classified as rare. In male mice the test of trend over dose was statistically significant ($p=0.0086 < 0.01$), as was the test in hepatocellular carcinoma and pooled adenoma and carcinoma ($p < 0.00005 < 0.01$). In female mice, results were even more highly statistically significant, for all three tests of trend in adenoma, carcinoma, and pooled adenoma and carcinoma (all $p < 0.00005 < 0.05$). In male mice the test comparing the high dose group to the vehicle control group in adenomas was not quite statistically significant ($p = 0.0171 > 0.01$). However, in both genders the tests comparing the high dose group to controls were highly statistically significant in both carcinomas and pooled adenomas and carcinomas ($p = 0.00005 < 0.01, 0.05$). The test comparing the high dose group to controls in adenomas in female mice was also highly statistically significant ($p = 0.0001 < 0.05$). In male rats the pairwise test of the medium dose group to the vehicle controls in adenoma was also statistically significant ($p = 0.0078 < 0.01$), as was the pairwise test in carcinoma and pooled adenoma and carcinoma (both $p < 0.00005 < 0.01$). No other pairwise comparison was even statistically significant at a nominal 0.05 level.

Table A.3.1. Potentially Statistically Significant Neoplasms in Mice

Tumor	Veh	Low	Med	High	Trend	Veh	Veh	Veh
						vs Low	vs Med	vs High
Male Mice Liver								
ADENOMA - BENIGN	12	18	26	23	0.0086	0.2653	0.0078	0.0171
HEPATOCELLULAR CARCINOMA	5	10	29	32	0.0000	0.2848	0.0000	0.0000
Adenoma/Hepato. Carc.	15	24	38	40	0.0000	0.1784	0.0000	0.0000

Table A.3.1.(cont.) Potentially Statistically Significant Neoplasms in Mice

Tumor					Trend	Veh	Veh	Veh
	Veh	Low	Med	High		vs Low	vs Med	vs High
Female Mice Liver								
ADENOMA - BENIGN	0	2	1	12	0.0000	0.2680	0.5143	0.0001
HEPATOCELLULAR CARCINOMA	0	0	2	15	0.0000	1.0000	0.2609	0.0000
Adenoma/Hepato. Carc.	0	2	2	24	0.0000	0.2680	0.2609	0.0000

Table A.3.2. Potentially Statistically Significant Neoplasms in Male Mice

Tumor					Trend	Veh	Veh	Veh
	Veh	Low	Med	High		vs Low	vs Med	vs High
Adrenals								
ADENOMA - BENIGN	0	0	0	1	0.2105	.	.	0.4667
PHAECHROMOCYTOMA - BENIGN	1	2	0	0	0.8731	0.5842	0.5152	0.4667
SUB-CASPULAR CELL ADENOMA	1	3	1	2	0.3576	0.3856	0.2615	0.4492
Bone (All)								
OSTEOSARCOMA - MALIGNANT	0	0	3	1	0.1553	.	0.1425	0.4667
Endothelium								
Hemangioma/-sarcoma	1	2	2	0	0.7529	0.5748	0.5231	0.4667
Hemangiosarcoma	1	2	2	0	0.7529	0.5748	0.5231	0.4667
Eyes								
AMELANOTIC MELANOMA - MALIGNANT	1	0	0	0	0.7594	0.5493	0.5152	0.4667
Fibrous Connective Tissue								
FIBROMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.
Fibroma/-sarcoma	1	2	1	1	0.4792	0.5748	0.2615	0.7198
Fibrosarcoma	1	1	1	1	0.4536	0.2982	0.2615	0.7198
Haemopoietic tissue								
FOLLICULAR CENTRE CELL LYMPHOMA	5	2	3	3	0.5148	0.8274	0.6452	0.5337
HISTIOCYTIC SARCOMA - MALIGNANT	2	0	1	1	0.4675	0.7934	0.5114	0.4376
SMALL LYMPHOCYTE LYMPHOMA	0	0	0	1	0.2164	.	.	0.4754
Harderian glands								
ADENOMA - BENIGN	7	3	9	6	0.2637	0.9120	0.5159	0.3956
Ileum								
ADENOCARCINOMA - MALIGNANT	0	1	0	0	0.4662	0.5493	.	.
Incisor teeth								
OSTEOSARCOMA - MALIGNANT	0	0	0	1	0.2105	.	.	0.4667
Liver								
ADENOMA - BENIGN	12	18	26	23	0.0086	0.2653	0.0078	0.0171
Adenoma/Hepato. Carc.	15	24	38	40	0.0000	0.1784	0.0000	0.0000
CHOLANGIOMA - BENIGN	1	0	0	0	0.7594	0.5493	0.5152	0.4667
HAEMANGIOMA - BENIGN	0	0	1	1	0.1515	.	0.5152	0.4667
HAEMANGIOSARCOMA - MALIGNANT	1	0	2	0	0.5836	0.5493	0.5231	0.4667
HEPATOCELLULAR CARCINOMA	5	10	29	32	0.0000	0.2848	0.0000	0.0000
Lungs								
ADENOCARCINOMA - MALIGNANT	6	5	7	4	0.5571	0.5990	0.5460	0.5257
ADENOMA - BENIGN	12	13	11	5	0.9351	0.5627	0.5977	0.8942
Mesenteric lymph nodes								
HAEMANGIOMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.
Oral cavity								
PAPPILOMA - BENIGN	1	0	0	0	0.7594	0.5493	0.5152	0.4667
Pancreas								
ISLET CELL ADENOMA - BENIGN	0	2	1	0	0.6324	0.2982	0.5224	.
Pituitary gland								
ADENOMA - BENIGN	0	1	2	0	0.4903	0.5493	0.2615	.
Prostate gland								
ADENOCARCINOMA - MALIGNANT	0	1	0	0	0.4662	0.5493	.	.
INTRATUBULAR ADENOMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.

Table A.3.2. (cont.) Potentially Statistically Significant Neoplasms in Male Mice

Tumor						Veh	Veh	Veh
	Veh	Low	Med	High	Trend	vs Low	vs Med	vs High
Seminal vesicles								
ANAPLASTIC CARCINOMA - MALIGNANT	0	0	1	0	0.4662	.	0.5152	.
Skin								
All Skin Tumors	4	7	2	1	0.9503	0.3987	0.6915	0.7675
HAEMANGIOMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.
HISTIOCYTOMA - BENIGN	1	0	0	0	0.7594	0.5493	0.5152	0.4667
KERATOACANTHOMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.
PAPILLOMA - BENIGN	1	0	0	0	0.7594	0.5493	0.5152	0.4667
RHABDOMYOSARCOMA - MALIGNANT	0	0	0	1	0.2105	.	.	0.4667
SPINDLE CELL SARCOMA - MALIGNANT	2	5	1	0	0.9615	0.3176	0.5231	0.7198
SQUAMOUS CELL CARCINOMA	0	0	1	0	0.4701	.	0.5224	.
Selected Skin Tumors	3	6	2	0	0.9762	0.3611	0.5284	0.8484
Spleen								
HAEMANGIOSARCOMA - MALIGNANT	0	2	0	0	0.7169	0.2982	.	.
Stomach								
PAPILLOMA - BENIGN	0	0	0	1	0.2105	.	.	0.4667
Subcutaneous fat								
CYSTADENOMA - BENIGN	0	1	0	0	0.4662	0.5493	.	.
Systemic								
Hemangioma	0	2	1	1	0.3514	0.3052	0.5152	0.4667
Testes								
INTERSTITIAL CELL ADENOMA	1	3	1	0	0.8567	0.3856	0.2615	0.4667
Thyroid glands								
FOLLICULAR ADENOMA - BENIGN	0	1	0	1	0.2759	0.5493	.	0.4667
Urinary bladder								
HISTIOCYTOMA - BENIGN	0	0	0	1	0.2105	.	.	0.4667
brown adipose tissue								
HIBERNOMA - BENIGN	1	0	1	0	0.5930	0.5493	0.2615	0.4667

Table A.3.3. Potentially Statistically Significant Neoplasms in Female Mice

Tumor						Veh	Veh	Veh
	Veh	Low	Med	High	Trend	vs Low	vs Med	vs High
Adrenals								
SUB-CASPULAR CELL ADENOMA	0	0	0	1	0.2357	.	.	0.4925
Bone								
Osteoma/-sarcoma	0	1	0	0	0.4929	0.5211	.	.
Broad ligament fat								
HAEMANGIOMA - BENIGN	0	1	0	0	0.4929	0.5211	.	.
Caecum								
LEIOMYOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Cervix								
LEIOMYOMA - BENIGN	0	0	0	1	0.2357	.	.	0.4925
Clitoral glands								
HISTIOCYTOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Cranium								
BASAL CELL CARCINOMA - MALIGNANT	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Endothelium								
HAEMANGIOMA - BENIGN	0	2	2	0	0.6427	0.2680	0.2609	.
Hemangioma/-sarcoma	0	4	3	1	0.5476	0.0717	0.1359	0.5000
Hemangiosarcoma	0	2	1	1	0.4053	0.2750	0.5143	0.5000
Eyes								
ADENOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Femur & joint (incl. marrow)								
OSTEOMA - BENIGN	0	1	0	0	0.4929	0.5211	.	.
Haemopoietic tissue								
FOLLICULAR CENTRE CELL LYMPHOMA	22	21	15	15	0.9090	0.6669	0.9229	0.8726
HISTIOCYTIC SARCOMA - MALIGNANT	1	5	4	4	0.2384	0.1269	0.2152	0.1873
SMALL LYMPHOCYTE LYMPHOMA	1	1	2	2	0.2404	0.2750	0.5217	0.4886

Table A.3.3. (cont.) Potentially Statistically Significant Neoplasms in Female Mice

Tumor					Trend	Veh	Veh	Veh
	Veh	Low	Med	High		vs Low	vs Med	vs High
Harderian glands								
ADENOCARCINOMA - MALIGNANT	0	1	0	0	0.4929	0.5211	.	.
ADENOMA - BENIGN	2	4	7	5	0.1592	0.3774	0.0893	0.2141
Kidneys								
CYSTADENOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Lacrimal glands								
ADENOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Liver								
ADENOMA - BENIGN	0	2	1	12	0.0000	0.2680	0.5143	0.0001
Adenoma/Hepato. Carc.	0	2	2	24	0.0000	0.2680	0.2609	0.0000
HAEMANGIOMA - BENIGN	1	0	0	1	0.4172	0.5211	0.5143	0.7463
HAEMANGIOSARCOMA - MALIGNANT	0	1	0	0	0.4894	0.5278	.	.
HEPATOCELLULAR CARCINOMA	0	0	2	15	0.0000	.	0.2609	0.0000
Lungs								
ADENOCARCINOMA - MALIGNANT	2	2	2	1	0.6700	0.3399	0.3290	0.4886
ADENOMA - BENIGN	7	8	6	1	0.9896	0.5695	0.5220	0.9667
Mesovarian fat								
LEIOMYOMA - BENIGN	0	1	0	0	0.4929	0.5211	.	.
Ovaries								
CYSTADENOMA - BENIGN	1	0	1	1	0.3812	0.5211	0.2609	0.7537
GRANULOSAR CELL - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
LUTEOMA - BENIGN	1	0	3	2	0.1673	0.5211	0.3290	0.5000
Pituitary gland								
ADENOMA - BENIGN	0	0	2	0	0.4819	.	0.2609	.
Rib								
OSTEOMA - BENIGN	0	1	0	0	0.4929	0.5211	.	.
Site of Mammary Gland								
ADENOCARCINOMA - MALIGNANT	1	1	2	1	0.4889	0.2680	0.5217	0.7537
Skin								
All Skin Tumors	2	3	1	0	0.9410	0.5535	0.5217	0.7463
RHABDOMYOSARCOMA - MALIGNANT	0	1	0	0	0.4929	0.5211	.	.
SPINDLE CELL SARCOMA - MALIGNANT	1	0	0	0	0.7571	0.5211	0.5143	0.4925
SQUAMOUS CELL CARCINOMA	0	2	1	0	0.6684	0.2750	0.5143	.
Selected Skin Tumors	1	2	1	0	0.8180	0.5422	0.2609	0.4925
Spleen								
HAEMANGIOMA - BENIGN	0	0	0	1	0.2357	.	.	0.4925
HAEMANGIOSARCOMA - MALIGNANT	0	1	0	0	0.4929	0.5211	.	.
HISTIOCYTIC SARCOMA - MALIGNANT	0	0	0	1	0.2411	.	.	0.5000
SMALL LYMPHOCYTIC LYMPHOMA	0	0	1	0	0.4965	.	0.5211	.
Stomach								
ADENOMA - BENIGN	0	1	0	0	0.4929	0.5211	.	.
PAPILLOMA - BENIGN	1	0	0	0	0.7571	0.5211	0.5143	0.4925
Systemic								
Hemangioma	1	3	2	2	0.4162	0.3399	0.5217	0.4886
Hemangioma/-sarcoma	1	5	3	3	0.3738	0.1269	0.3399	0.3068
Hemangiosarcoma	0	2	1	1	0.4053	0.2750	0.5143	0.5000
Thyroid glands								
FOLLICULAR ADENOMA - BENIGN	0	0	1	2	0.0572	.	0.5143	0.2463
Uterus								
ADENOCARCINOMA - MALIGNANT	0	1	0	0	0.4929	0.5211	.	.
HAEMANGIOMA - BENIGN	0	2	2	0	0.6427	0.2680	0.2609	.
HAEMANGIOSARCOMA - MALIGNANT	0	0	1	1	0.1809	.	0.5143	0.5000
HISTIOCYTOMA - BENIGN	0	0	0	2	0.0568	.	.	0.2463
LEIOMYOMA - BENIGN	0	2	0	0	0.7446	0.2680	.	.
LEIOMYOSARCOMA - MALIGNANT	0	0	1	1	0.1809	.	0.5143	0.5000
Leiomyoma/-sarcoma	0	2	1	1	0.4086	0.2680	0.5143	0.5000
POLYP - BENIGN	3	2	3	1	0.7767	0.5268	0.3640	0.6820

Appendix 4. Bayesian Tumorigenicity Analysis

The frequentist approach to testing in the presence of multiplicities is to adjust the type I error rate (i.e., the probability of rejecting a true hypothesis of no differences). For example, the Haseman-Lin-Rahman rules, described in Section 1.3.1.4, are designed to control the type I error for tests of trend and for pairwise tests each at about a 10% error rate and apply to Peto tests, and apparently also the poly-k tests. The Bayesian approach is not tied to Type I error, and assesses the probability of each of the multiple events on the basis of all information in the trial, including other events. Unlike the frequentist analysis, the fact that the Bayesian assessments are conditional on observed data allows one to specify analyses conditional on data based criteria. The criterion used here was that there should be at least one tumor in the high dose group and one or more tumors in the remaining dose groups.

For this analysis we define a two stage mixture model on the treatment parameters in a simple logit model for tests of trend and pairwise comparisons. For testing trend, we define p_{ijk} as the probability of tumor i in subject j in treatment group k . That is, with $i = 1$ to n_t tumors and $j = 1$ to n_s subjects (here $n_s = 54$), and dose d_k , (here $k=1-4$), leaving the experiment at time t_j either from death or sacrifice, and subject effect δ_j :

$\text{logit}(p_{ijk}) = \alpha_i + \beta_i d_k + \gamma_i t_j + \delta_j$, $k=1, \dots, 4$, $i=1, \dots, n_t$, $j=1, \dots, n_s$,
with random subject effect $\delta_j \sim N(\mu_\delta, \sigma_\delta^2)$. We assign model priors:

$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$
 $\beta_i \sim \pi_i I_{[0]} + (1 - \pi_i)N(\mu_\beta, \sigma_\beta^2)$ for $i=1, \dots, n_t$
 and, $\pi_i \sim \text{Beta}(1,3)$,
 $\gamma_i \sim N(\mu_g, \sigma_g^2)$ for $j=1, \dots, n_s$.
 with $\mu_\delta = \mu_\alpha = \mu_\beta = \mu_g = \mu_s = 0$ and $\sigma_\delta^2 = 100$,
 $\sigma_\alpha^2, \sigma_\beta^2, \sigma_g^2 \sim \text{Inverse Gamma}(1,3)$.

The model for pairwise comparisons is similar:

$\text{logit}(p_{ijk}) = \alpha_i + \beta_{ik} + \gamma_i t_j + \delta_j$, $k=2,3,4$, $i=1, \dots, n_t$, $j=1, \dots, n_s$, with $\beta_{i0} = 0$,
though here only $k = 4$ (i.e. high dose) was investigated.

$\pi_{ik} \sim \text{Beta}(1,3)$ for $k = 2,3,4$.

$\beta_{ik} \sim \pi_{ik} I_{[0]} + (1 - \pi_{ik})N(\mu_{\beta k}, \sigma_{\beta k}^2)$ for $i=1, \dots, n_t$, $k = 2,3,4$.

Note that with this parameterization, for $k = 2,3,4$, the β_{ik} represent the deviation of treatment effect from the controls.

These are meant to reflect uncertainty about the values of the parameters. Initial ignorance about the exact values of the α , γ , and δ parameters are modeled with well dispersed normal distributions centered at zero. The β parameters representing the either effect of linear effect dose or difference between the high dose and control are weighted so that they initially are weighted towards 0. Thus, if the posterior probability they are zero is very small, this provides strong evidence that the parameter is not zero. In the Bayesian vernacular, the priors above should be reasonably noninformative about the parameters.

As noted before, the choice of tumors chosen for analysis is conditioned on there being at least one tumor in the high dose group and at least two or more tumors overall. Tables A.5.1 and A.5.2 below indicate the observed frequency of tumors and the estimated posterior probability that the linear dose effect (i.e., slope) is zero. The rightmost column is the estimated posterior probability that the differential effect of the high dose over the control effect is zero.

Table A.4.1 Incidence of Tumors in Males Used in Bayesian Analysis

#	Organ	Tumor	cntrl	low	med	high	Probabilities	
							slope = 0	High=Cntrl
1	Fibrous Connective Tissue	Fibroma/-sarcoma	1	2	1	1	0.973	0.8744
2	Fibrous Connective Tissue	Fibrosarcoma	1	1	1	1	0.9747	0.8752
3	Haemopoietic tissue	Any Lymphoma	5	2	3	4	0.9838	0.8907
4	Haemopoietic tissue	FOLLICULAR CENTRE CELL LYMPHOMA	5	2	3	3	0.9696	0.8537
5	Haemopoietic tissue	HISTIOCYTIC SARCOMA - MALIGNANT	2	0	1	1	0.9864	0.9205
6	Liver	ADENOMA - BENIGN	12	18	26	23	0.7864	0.4777
7	Liver	Adenoma/Hepato. Carc.	17	28	55	55	0.0	0.0
8	Liver	HAEMANGIOMA - BENIGN	0	0	1	1	0.9674	0.849
9	Liver	HEPATOCELLULAR CARCINOMA	5	10	29	32	0.0	0.0
10	Lungs	ADENOCARCINOMA - MALIGNANT	6	5	7	4	0.9874	0.9182
11	Lungs	ADENOMA - BENIGN	12	13	11	5	0.9404	0.7755
12	Systemic	Hemangioma	0	2	1	1	0.9767	0.847
13	Systemic	Hemangioma/-sarcoma	1	4	3	1	0.976	0.8718
14	Adrenals	SUB-CASPULAR CELL ADENOMA	1	3	1	2	0.9897	0.9448
15	Bone (All)	OSTEOSARCOMA - MALIGNANT	0	0	3	1	0.9729	0.8511
16	Harderian glands	ADENOMA - BENIGN	7	3	9	6	0.981	0.8917
17	Skin	All Skin Tumors	4	7	2	1	0.9742	0.8503
18	Thyroid glands	FOLLICULAR ADENOMA - BENIGN	0	1	0	1	0.8566	0.7424

Thus in males there is strong evidence of a dose related trend in hepatocellular carcinoma and pooled adenoma and hepatocellular carcinoma in the liver, as well as the comparison between the high dose and control for both tumor types (i.e., for p =probability of trend or difference between high and control, all $p \leq 0.00005$). Over all four dose groups, the distribution of the slope parameter is concentrated on the interval 0.36 to 0.69 for pooled adenoma and hepatocellular carcinoma and in 0.37 to 0.73 for hepatocellular carcinoma (i.e. these intervals are estimated 95% credible intervals). The corresponding 95% credible intervals for the differences in incidence are 1.96 to 4.29 and 2.04 to 4.31, respectively. The simple probability of a difference between the high dose and controls in adenomas in the liver is 0.5223 (= 1-0.4777).

It should be emphasized that the probabilities cited above are not significance levels. In a frequentist analysis, as exemplified in the poly-k analysis in Appendix 3, the significance levels are the estimated probability of a result as “extreme” or more extreme than that which was observed, assuming there is no effect of dose. In the Bayesian approach, if we assume the priors represent reasonable expressions of uncertainty about the parameters, then the probabilities cited above represent reasonable expressions of the probability that the parameter is actually zero. Both approaches assume the probability models are reasonable depictions of reality.

Table A.4.2 Incidence of Tumors in Females Used in Bayesian Analysis

#	Organ	Tumor	cntrl	low	med	high	Probabilities	
							slope =0	High=Cntrl
1	Haemopoietic tissue	Any Lymphoma	23	22	18	17	0.9755	0.8829
2	Haemopoietic tissue	FOLLICULAR CENTRE CELL LYMPHOMA	22	21	15	15	0.9843	0.8397
3	Haemopoietic tissue	HISTIOCYTIC SARCOMA - MALIGNANT	1	5	4	4	0.9803	0.8936
4	Haemopoietic tissue	SMALL LYMPHOCYTE LYMPHOMA	1	1	2	2	0.9844	0.9127
5	Liver	ADENOMA - BENIGN	0	2	1	12	0.011	0.00988
6	Liver	Adenoma/Hepato. Carc.	0	2	3	27	0.0	0.00204
7	Liver	HAEMANGIOMA - BENIGN	1	0	0	1	0.9724	0.8795
8	Liver	HEPATOCELLULAR CARCINOMA	0	0	2	15	0.0	0.0
9	Lungs	ADENOCARCINOMA - MALIGNANT	2	2	2	1	0.9744	0.8581
10	Lungs	ADENOMA - BENIGN	7	8	6	1	0.7373	0.3762
11	Ovaries	CYSTADENOMA - BENIGN	1	0	1	1	0.9758	0.8947
12	Ovaries	LUTEOMA - BENIGN	1	0	3	2	0.9745	0.873
13	Systemic	Hemangioma	1	3	2	2	0.986	0.8875
14	Systemic	Hemangioma/-sarcoma	1	5	3	3	0.9784	0.8462
15	Systemic	Hemangiosarcoma	0	2	1	1	0.9873	0.878
16	Uterus	HAEMANGIOSARCOMA - MALIGNANT	0	0	1	1	0.883	0.8196
17	Uterus	HISTIOCYTOMA - BENIGN	0	0	0	2	0.9679	0.8028
18	Uterus	LEIOMYOSARCOMA - MALIGNANT	0	0	1	1	0.9779	0.8445
19	Uterus	Leiomyoma/-sarcoma	0	2	1	1	0.9625	0.8421
20	Uterus	POLYP - BENIGN	3	2	3	1	0.9634	0.8421
21	Harderian glands	ADENOMA - BENIGN	2	4	7	5	0.982	0.8718
22	Site of Mammary Gland	ADENOCARCINOMA - MALIGNANT	1	1	2	1	0.9419	0.8229
23	Thyroid glands	FOLLICULAR ADENOMA - BENIGN	0	0	1	2	0.9781	0.8753

Thus, in females, as with males there is strong evidence of a dose related trend in hepatocellular carcinoma and pooled adenoma and hepatocellular carcinoma in the liver (i.e., for p =probability of trend, both $p \leq 0.00005$). The same comment holds for the difference between the high dose group and controls for hepatocellular carcinoma in the liver. For pooled adenoma and hepatocellular carcinoma in the liver the probability of a difference between the high dose group and controls is 0.002. Over all four dose groups, the distribution of the slope parameter is concentrated on the interval 0.46 to 1.12 for pooled adenoma and hepatocellular carcinoma and in 0.56 to 1.06 for hepatocellular carcinoma (i.e. these intervals are estimated 95% credible intervals). The corresponding 95% credible intervals for the differences in incidence are 2.4 to 6.24 and 1.61 to 5.85, respectively. The posterior probability that the slope is zero in simple adenomas in the liver is 0.011, while the probability that the difference between the high dose and control is zero is 0.0099. The corresponding 95% credible intervals are 0.28 to 0.86 and 1.16 to 5.5, respectively.

These analyses were implemented in WinBUGS 1.4.3 (see Lunn et al, 2000)

Appendix 5. 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.
- Haseman, J. K. (1983), A Reexamination of False-positive Rates for Carcinogenicity Studies, *Fundamental and Applied Toxicology*, 3, 334-339.
- Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001), *Bayesian Survival Analysis*, Springer Verlag, New York.
- 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.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325--337.
- 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.
- 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.
- 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.

SAS Institute (2008), *SAS/STAT[®] 9.2 User's Guide* The MCMC Procedure (Experimental), Cary, North Carolina (available online), 3475-3697.

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

U.S. Department of Health and Human Services (2001), *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.

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

/s/

Steven Thomson
7/22/2009 10:04:16 AM
BIOMETRICS

Karl Lin
7/22/2009 10:06:16 AM
BIOMETRICS
Concur with review