CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

203202Orig1s000

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

ADDENDUM

NDA/BLA Serial Number:	NDA 203-202 (SN 0048)
Drug Name:	Droxidopa
Indication(s):	treatment of symptomatic neurogenic orthostatic hypotension (NOH) in adult patients with primary autonomic failure [Parkinson's Disease (PD), Multiple System Atrophy (MSA) and Pure Autonomic Failure (PAF)], Dopamine Beta Hydroxylase (DβH) Deficiency and Non-Diabetic Autonomic Neuropathy (NDAN)
Applicant:	Chelsea Therapeutics, Inc
Date(s):	Date of Document: August 13, 2013 PDUFA due date: February 13, 2014
Review Priority:	Priority
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	James Hung, Ph.D.
Medical Division:	Division of Cardiovascular and Renal Products, HFD-110
Clinical Team:	Shari Targum, M.D.
Project Manager:	Anna Park, Pharm.D.

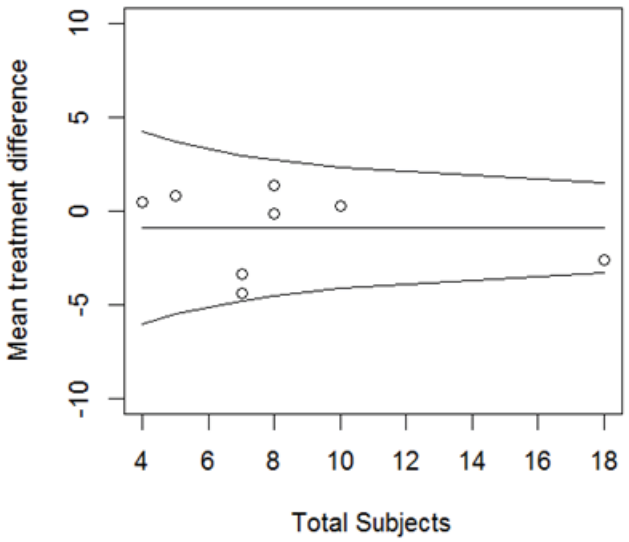
Keywords:

Intra-subject variability, funnel plot

This addendum provides additional analyses and results that were not included in the original statistical review of the resubmission.

Site 132

Site 132 is the largest site in Study 306B. The site enrolled a total of 25 patients. 5 patients were unblinded for the interim analysis and therefore were excluded from Study 306B. The rest of the patients (20) were included in study 306B. Two of the 20 patients discontinued before Week 1 and therefore were excluded from the full analysis set for the primary analysis. The mean treatment effect in that site was 2.6 units in OHSA Item 1. The site was selected for inspection due to its large treatment effect and number of enrollment. No violations were discovered by the FDA site inspection. By excluding the site, the overall treatment effect decreased from 0.94 unit in OHSA Item 1 to 0.68 unit and the p-value changed from 0.028 to 0.13. Funnel plot did not suggest that the treatment effect in this site, although large, was clearly an outlier. Although removal of Site 132 changed the primary analysis result from "statistically significant" to "statistically non-significant" (i.e., p-value from 0.028 to 0.13), this reviewer did not find any compelling reason to exclude this site from the overall population.



Intra-subject variability

The reviewer calculated the intra-subject variability based on the post-baseline OHSA Item 1 assessments. The value reported in the original statistical review was 2.9 which was the variance of the measurements. The standard deviation was 1.7 (i.e., square root of 2.9). The standard deviation was presented to the Advisory Committee Meeting on January 14, 2014.

The original review mentioned that the treatment effect of 0.9 unit was small when compared with the intra-subject variability. The reviewer further explored the concept with additional calculations. For example, given the intra-subject variability with a standard deviation of 1.7 unit, how likely would we observe that a typical subject has a 0.9 unit or greater difference in OHSA Item 1 between two consecutive visits without any treatment?

First, the placebo data were used to estimate directly the number of cases where the differences in OHSA Item 1 measurements between consecutive visits exceeded 0.9 unit. The total number of post-baseline visits that had a difference of at least 0.9 unit from the previous visit in the placebo group were counted. This number was then divided by the total number of post-baseline visits. Missing visits were excluded in the estimation. The result is that approximately 70% post-baseline visits had a difference exceeding 0.9 unit from the previous visit. Secondly, the reviewer computed the probability that two independent normally distributed random variables X and Y, each with a standard deviation of 1.7, differ by at least 0.9 unit; the probability P(|X-Y|>0.9) is 0.71. Based on the placebo group data, the correlation between two post-baseline consecutive visits was approximately 0.5 to 0.6. Using bivariate normal distribution with correlation of 0.5 for calculation, the probability that two consecutive visits have a difference of at least 0.9 unit is approximately 0.6. These calculations give some insight into how likely we would see a difference of at least 0.9 unit between two consecutive visits in the placebo group.

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

/s/

JIALU ZHANG
02/18/2014

HSIEN MING J HUNG
02/18/2014

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA Serial Number:	NDA 203-202 (SN 0048)
Drug Name:	Droxidopa
Indication(s):	treatment of symptomatic neurogenic orthostatic hypotension (NOH) in adult patients with primary autonomic failure [Parkinson's Disease (PD), Multiple System Atrophy (MSA) and Pure Autonomic Failure (PAF)], Dopamine Beta Hydroxylase (DβH) Deficiency and Non-Diabetic Autonomic Neuropathy (NDAN)
Applicant:	Chelsea Therapeutics, Inc
Date(s):	Date of Document: August 13, 2013 PDUFA due date: February 13, 2014
Review Priority:	Priority
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	James Hung, Ph.D.
Medical Division:	Division of Cardiovascular and Renal Products, HFD-110
Clinical Team:	Shari Targum, M.D.
Project Manager:	Anna Park, Pharm.D.

Keywords:

Neurogenic Orthostatic Hypotension, dropout, ANCOVA

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

The original NDA 203202 was submitted on September 28, 2012 by the sponsor to seek approval of droxidopa in treating symptomatic Neurogenic Orthostatic Hypotension (NOH) associated with Parkinson's disease (PD), Multiple System Atrophy (MSA), Pure Autonomic Failure (PAF), Dopamine Beta Hydroxylase (DBH) deficiency, or Non-Diabetic Autonomic Neuropathy (NDAN). This NDA resubmission included Study 306B to address the deficiencies listed in the Complete Response Letter issued on March 28, 2012.

Study 306B was a multi-center, randomized, parallel-group, placebo-controlled, double-blind study with an initial dose titration, followed by an 8-week treatment period to evaluate the clinical effects of droxidopa in patients with symptomatic Neurogenic Orthostatic Hypotension (NOH) associated with Parkinson's Disease (PD).

After changing the primary endpoint twice, the final primary efficacy endpoint was the mean change in the OHSA Item 1 from Baseline to Week 1. The droxidopa group had a treatment effect of -0.94 compared to the placebo group in the change of OHSA Item 1 score from Baseline to Week 1. The p-value was 0.028 based on ANCOVA model and was statistically significant. Other measurements at Week 1, such as OHQ composite score, clinician and patient reported CGI-I and CGI-S, and standing systolic blood pressure (SBP) were all trending in the right direction, though might not reach statistical significance.

Study 306 went through a number of major changes during its course of conduct including changing the primary endpoint twice, splitting into Study 306A and Study 306B, and changing the total sample size. In addition, it was discovered that the unblinded statistical team had access to the treatment codes for all Study 306 subjects rather than the 51 patients for the interim analysis. Although the access was later revoked, a considerable number of patients in Study 306 were already enrolled. In order to address the concerns on study conduct, the sponsor performed a post-interim sensitivity analysis to show that the study results remained consistent. The reviewer also performed similar analyses at additional time points, such as after revoking access to treatment code and after changing to the final primary endpoint. The treatment effects in various measurements were all trending in the right direction but the magnitude of the treatment effect tended to be less for the patients who enrolled later during the trial.

Although the primary endpoint was statistically significant, the treatment effect on OHSA Item 1 at Week 1 seemed small at the presence of 2.9 unit of intra-subject variability. Also it is questionable whether droxidopa has any long term treatment effect. This was reflected in the diminishing treatment effect on OHSA Item 1 as well as standing SBP in later weeks in the study.

In addition, the imbalance of dropouts between droxidopa group and placebo group was concerning. 20 droxidopa patients were excluded from the primary analysis compared with only 7 placebo patients. Except for three untreated patients, the rest of these patients dropped out early in the study and had missing OHSA Item 1 score at Week 1. Even if excluding 8 patients who

enrolled earlier before the interim analysis, Study 306B still had 4 patients treated with placebo and 12 patients treated with droxidopa discontinued study prior to Week 1. The imbalance remained. The treatment effect on OHSA Item 1 became -0.45 with 95% confidence interval (-1.2, 0.3) if missing data were imputed by carrying forward the baseline observation (BOCF).

2. INTRODUCTION

2.1 Overview

This NDA resubmission included a single phase 3 trial Study 306B. Study 306B was a multi-center, randomized, parallel-group, placebo-controlled, double-blind study with an initial dose titration, followed by an 8-week treatment period to evaluate the clinical effects of droxidopa in patients with symptomatic Neurogenic Orthostatic Hypotension (NOH) associated with Parkinson's Disease (PD).

The trial started as Study 306 and had one interim analysis planned at N=50. The interim analysis was performed on 51 patients who completed 8 weeks of treatment. The DMC recommended terminating the trial due to futility following this interim analysis. After a period of reconsideration, the sponsor decided to continue the study but split the study into Study 306A (which contained 51 unblinded patients used for interim analysis) and Study 306B. The primary endpoint was also changed from OHQ composite score at Week 8 to patient-reported falls at Week 8. The primary endpoint was changed again from patient-reported falls at Week 8 to OHSA 1 at Week 1 after the original NDA was submitted. By then, 122 patients were randomized in Study 306B. Table 1 summarized the two studies included in the NDA resubmission.

Table 1. Efficacy Studies in the NDA Resubmission

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
306B	*Phase 3*	*Up to 2 week titration and 8 weeks of treatment*	*2 weeks*	*85 in placebo and 89 in droxidopa arm*	*Parkinson's Disease*
306A		*Up to 2 week titration and 8 weeks of treatment*	*2 weeks*	*27 in placebo and 24 in droxidopa arm*	*Parkinson's Disease*

The original NDA included three efficacy trials. The pivotal Study 301 was an induction-design trial with a 7-day double-blind randomized treatment period after an open-label dose-titration period and a washout period. The supportive Study 302 was a randomized withdrawal trial with 14-day double-blind randomized withdrawal period. Study 303 was designed to evaluate long-term safety and efficacy of droxidopa by a three-month open-label treatment period followed

with a double-blind randomized withdrawal phase. The NDA was submitted on September 28, 2011. The Division issued a complete response letter on March 28, 2012 stating that "the results of studies 302 and 303 undercut the persuasiveness of study 301" and "the disproportionate contribution of Site 507 to the overall results of study 301 diminishes the persuasiveness of the study". Please refer to the statistical reviews filed in December 2011 and March 2012 for further details.

2.2 Data Sources

The derived analysis datasets and raw datasets for Study 306B can be found under directory \\CDSESUB1\evsprod\NDA203202\0048\m5\datasets\noh306b.

The derived analysis datasets and raw datasets for Study 306A can be found under directory \\CDSESUB1\evsprod\NDA203202\0048\m5\datasets\noh306a

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

This NDA resubmission (SN0044) was first submitted on July 3, 2013 and had a number of data-related issues. The division issued an Incomplete Response Letter on July 29, 2013 listing all the data deficiencies, for example, the definition file for 12 raw datasets was missing, and the variable names in analysis datasets used for primary and secondary analyses did not match the variable names in the definition file. The NDA was resubmitted on August 13, 2013. To address the inconsistency of variable names between the datasets and the definition file, the sponsor created new definition files by adding a column with all variable names in the datasets and remapping them to the names in the old definition file. The datasets remained unchanged. However, this did not address the inconsistency of variable names between the SAS programs and the datasets. The so-claimed fully executable programs were not executable due to the inconsistency of variable names.

Nevertheless, the reviewer managed to trace how the primary endpoint was derived. The reviewer was also able to derive same or similar results in most of the primary and secondary analyses results from the CRF raw datasets submitted by the sponsor.

3.2 Evaluation of Efficacy

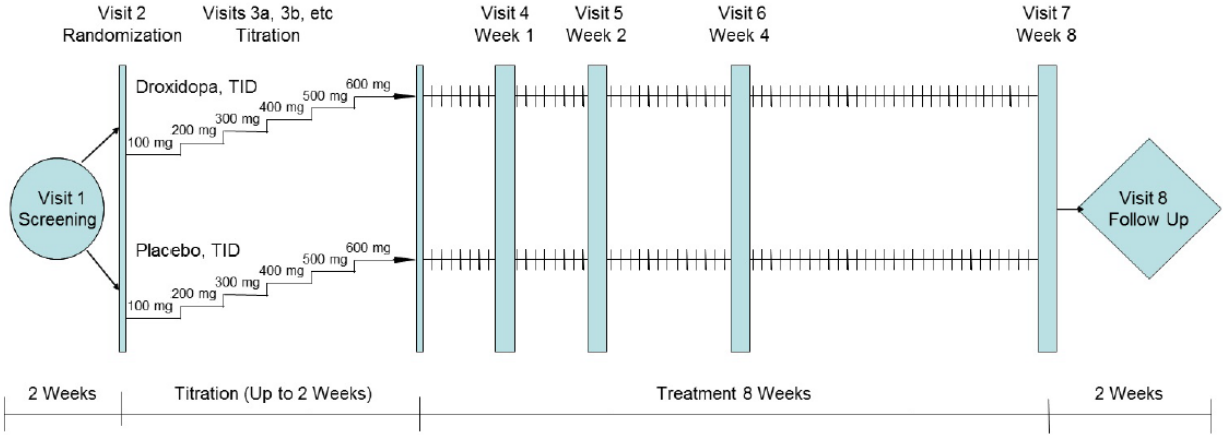
Study Design and Endpoints

Study 306B was a randomized, parallel-group, placebo-controlled, double-blind study with an initial dose titration (up to 14 days), followed by an 8-week treatment period (**Figure 1**). Patients were randomized in a ratio of 1:1 to either droxidopa or placebo at the end of Baseline Visit. Patients then had up to 14 days of double-blind titration starting at 100mg three times daily (TID) of droxidopa or matching placebo. Treatment was escalated in 100 mg TID increments until one of the titration stopping rules was met:

1. The patient became completely asymptomatic for NOH symptoms (clinician-reported CGI-S=1). At the Investigator's discretion, dose escalation may have been stopped when a patient became nearly asymptomatic (clinician-reported CGI-S=2)
2. The patient had a SBP>=180 mmHg or DBP>=110 mmHg after 10 minutes in supine position. At the Investigator's discretion, dose escalation may have been stopped when a patient's BP was close to the limits and further escalation was likely to result in BP levels exceeding the acceptable limit.
3. The patient was unable to tolerate side effects
4. The patient reached a maximum dose of 600 mg TID

Patients who met criterion 1 directly proceeded to the 8-week double-blind treatment period at that dose. Patients who met criterion 2 or 3 proceeded directly to the 8-week treatment period at the previous lower dose. Patients who met criterion 2 or 3 at initial dose of 100 mg TID were withdrawn from treatment. Patient who met criterion 4 continued into the 8-week treatment at 600 mg TID.

Figure 1. Study Design



[Source: Sponsor's clinical study report Figure 9-1]

The primary efficacy endpoint in Study 306B was the mean change in the OHSA Item 1 score from Baseline to Week 1. **The primary endpoint was changed twice during the trial**. The trial started as Study 306 and it was designed to measure the long-term safety and efficacy of droxidopa. The original primary endpoint was the mean change in OHQ composite score from Baseline to Week 8. An interim analysis was planned to assess study sample size at N=50. The actual interim analysis was performed when 51 patients completed 8-week treatment. The DMC recommended terminating the trial due to futility in January 2011 following the interim analysis. After a period of reconsideration, the sponsor decided to continue the study and changed the primary endpoint to patient-reported falls at Week 8. To maintain study integrity, the study was split to Study 306A (which contained 51 unblinded patients used for interim analysis) and Study 306B. The primary endpoint was changed again from patient-reported falls to OHSA Item 1 score at Week 1 in November 2011. The change was reflected in protocol version 4 dated November 5, 2011. By then, 122 patients were randomized in Study 306B. A total of 174 patients were enrolled into Study 306B and the last patient enrolled on August 10, 2012.

The sponsor planned to have 200 patients (100 patients each arm) when the primary endpoint was patient-reported falls. According to protocol version 4, this would provide 80% power to detect a treatment difference of 0.5 in patient-reported falls. The decision on terminating the study was announced in July 2012 and the total number of patients enrolled in the study was 174 (85 in placebo and 89 in droxidopa). The sponsor claimed that the trial was prematurely stopped based in FDA Advice Letter dated June 29, 2012. The letter expressed concerns that it was "not possible to know with certainty that interim results did not somehow influence decisions to change the primary efficacy endpoint of study 306". On the other hand, it was not clear to the reviewer whether the sponsor intended to keep the same sample size after changing the primary endpoint to OHSA Item 1 at week 1. The only protocol that reflect the change on the final primary endpoint OHSA Item 1 (version 5) was dated on November 2, 2012, which was after the last patient completed the study (October 23, 2012). The final SAP was dated on October 4, 2012 and was also after the enrollment was stopped.

The secondary efficacy variables in Study 306B were:

- The mean change in OHSA Item #1 from Baseline to week 2 (Visit 5)
- The mean change in OHSA Item #1 from Baseline to week 4 (Visit 6)
- The mean change in the lowest standing systolic blood pressure between 0 and +3 minutes of standing from Baseline to week 1 (Visit 4)
- The mean change in OHSA Item #1 from Baseline to week 8 (Visit 7)
- Rate of patient reported falls from Baseline to the end of the study (FAS)
- The mean change in OHQ from Baseline to week 8 (Visit 7)

The secondary endpoints were tested sequentially in the order listed above if the primary efficacy endpoint won at significance level of 0.05.

Patient Disposition, Demographic and Baseline Characteristics

A total of 174 patients were randomized in Study 306B (89 patients in droxidopa and 85 patients in placebo). 28% droxidopa patients discontinued study early compared to 20% placebo patients (**Table 2**).

Table 2. Patient Disposition

	Placebo (N=85) n (%)	Droxidopa (N=89) n (%)	Total (n=174) n (%)
Total Patients Randomized	85 (100)	89 (100)	174 (100)
Total Patients Treated	84 (98.8)	87 (97.8)	171 (98.3)
Completed Study	67 (78.8)	62 (69.7)	129 (74.1)
Discontinued Study	17 (20.0)	25 (28.1)	42 (24.1)
Reason for Discontinuation			
Treatment Failure	1 (1.2)	1 (1.1)	2 (1.1)
Adverse Event	6 (7.1)	10 (11.2)	16 (9.2)
Lack of Efficacy	2 (2.4)	4 (4.5)	6 (3.4)
Protocol Violation	0	1 (1.1)	1 (0.6)
Lost to Follow Up	1 (1.2)	0	1 (0.6)
Patient Withdrew Consent	1 (1.2)	3 (3.4)	4 (2.3)
Investigator Decision	1 (1.2)	2 (2.2)	3 (1.7)
Other	5 (5.9)	4 (4.5)	9 (5.2)

[Source: Sponsor's Clinical Study Report Table 10-1, verified by the reviewer]

Table 3 listed the three analysis populations. The Safety Set consisted of all patients who received at least one dose of study drug. The Full Analysis Set (FAS) was the population used for the primary analysis and consisted of all randomized patients who received at least one dose of study drug and reported OHSA Item 1 data at Week 1. Only 69 patients in droxidopa group were included in the primary analysis compared to 78 patients in placebo.

The Per Protocol Set consisted of patients in the FAS who were compliant with study treatment. Patients must have taken at least 80% of their planned study drug during the first four weeks of the treatment period and during the final four weeks of the treatment period.

Table 3. Analysis Populations

	Placebo (N=85)	Droxidopa (N=89)	Total (n=174)
Analysis Populations			
Safety Set	82 (96.5)	89 (100.0)	171 (98.3)
Full Analysis Set	78 (91.8)	69 (77.5)	147 (84.5)
Per Protocol Set	45 (52.9)	34 (38.2)	79 (45.4)

[Source: Sponsor's Clinical Study Report Table 11-1, verified by the reviewer]

The majority of patients in both treatment groups were male (69.7% in droxidopa group and 63.4% in placebo group). The mean ages were 72.5 years and 72.0 years for patients in the droxidopa and placebo groups, respectively. Most patients were White (95.5% in droxidopa group and 96.3% in placebo group). All patients were enrolled in the US.

Table 4. Demographic and Baseline Characteristics (Safety Set)

	Placebo (N=82)	Droxidopa (N=89)
Sex [n (%)]		
Male	52 (63.4)	62 (69.7)
Female	30 (36.6)	27 (30.3)
Race [n (%)]		
White	79 (96.3)	85 (95.5)
Black/African American	1 (1.2)	2 (2.2)
Asian	0	1 (1.1)
Hispanic/Latino	2 (2.4)	1 (1.1)
Age (Years) at Screening		
Mean (SD)	72.01 (8.036)	72.54 (7.571)
Min, Max	52.9, 86.3	41.4, 91.7
Weight (kg)		
Mean (SD)	77.06 (15.913)	78.03 (17.002)
Min, Max	45.5, 122.3	46.4, 122.0

[Source: Sponsor's clinical study report Table 11-2, verified by the reviewer]

Statistical Methodologies

The primary efficacy analysis was based on the Full Analysis Set. According to the sponsor's final SAP, the primary endpoint would be tested using analysis of covariance (ANCOVA) model adjusting for Baseline OHSA Item 1 score. However, if any of the ANCOVA assumptions (independence, constant variance or normality of the residuals) were not met then the primary analysis would be changed to non-parametric model using rank statistics adjusted for the OHSA Item 1 at Baseline. The violation of assumptions was determined by visually inspecting the diagnostic plots and no formal test was proposed.

The analysis of patient-reported falls was performed for all subjects' data in the FAS and included all data while subjects were in the study. The other secondary efficacy endpoints were analyzed with missing data excluded. LOCF was used as a sensitivity analysis.

Results and Conclusions

The sponsor reported that the assumptions for the ANCOVA were not met and used non-parametric methodology instead for the primary analysis. The resulting p-value was 0.018 and the treatment difference in OHSA Item 1 score was -1.0 with 95% confidence interval (-2.0, 0). The reviewer, however, did not find any obvious deviation from ANCOVA assumptions. Table 5 summarized the reviewer's results on primary endpoint by ANCOVA. The droxidopa group had a treatment effect of -0.94 when compared to placebo group in terms of change in OHSA Item 1 score from Baseline to Week 1. The p-value was 0.028 and was statistically significant.

Table 5. Primary Endpoint Results

	Droxidopa			Placebo		
	N	Mean	STD	N	Mean	STD
Baseline	69	5.1	2.04	78	5.1	2.33
Week 1	69	2.8	2.44	78	3.8	2.75
Least square mean difference	-0.94 with 95% CI (-1.78, -0.1)					
p-value from ANCOVA model	0.028					

Figure 2 showed the cumulative distribution of the change in the OHSA Item 1 score from Baseline to Week 1. **Figure 3** displayed the relationship between the baseline OHSA Item 1 score and the change in the OHSA Item 1 score from Baseline to Week 1. The two parallel lines are the estimated values of the change in OHSA 1 from Baseline in placebo group (blue) and in droxidopa group (grey) from the ANCOVA model in the primary analysis. The magnitude of change in the OHSA Item 1 from Baseline to Week 1 had a strong linear relationship with the baseline OHSA Item 1. The variability also seemed large. The intra-subject variability was 2.9. The reviewer calculated the intra-subject variability by including only the post-baseline visits (Week 1, Week 2, Week 4 and Week 8). Although the treatment effect on OHSA Item 1 at Week 1 reached statistical significance, the magnitude of the treatment effect (1 unit) seemed small when compared to the intra-subject variability.

Figure 2. Cumulative Distribution on the Change of OHSA Item 1 from Baseline at Week 1

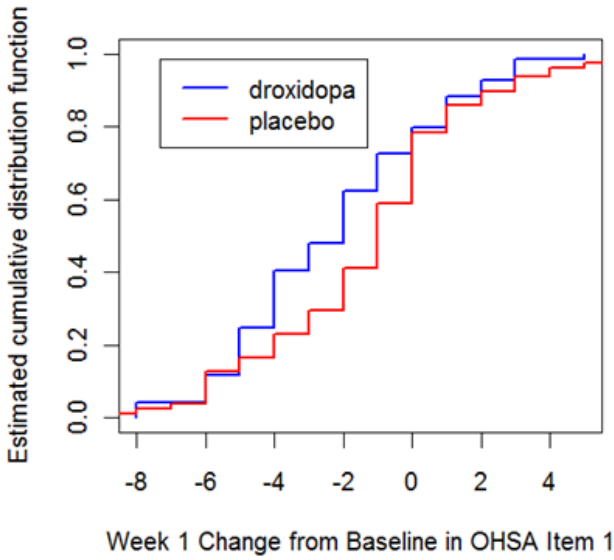
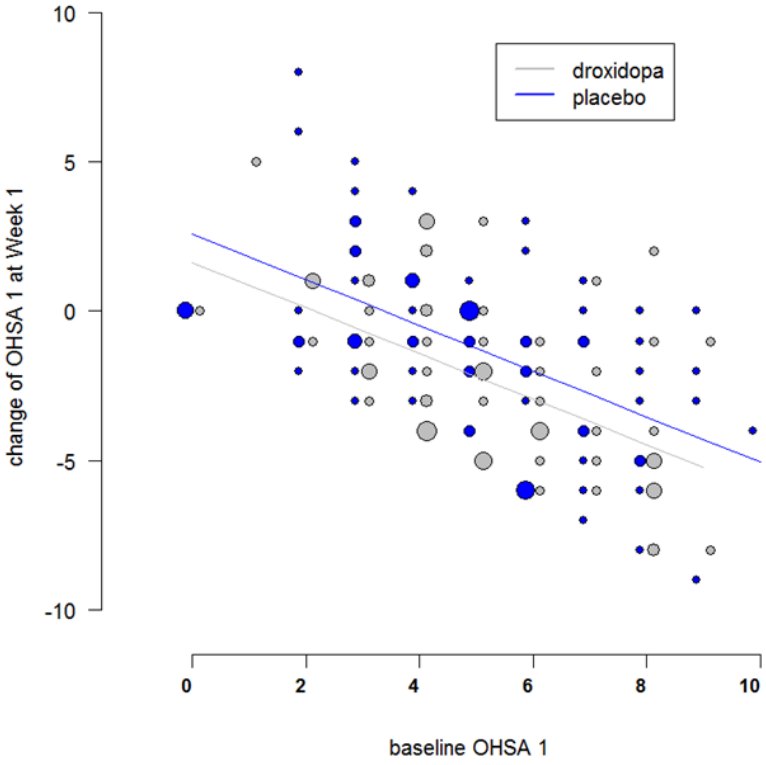


Figure 3. Change of OHSA 1 at Week 1 versus Baseline OHSA 1



* bigger circle represents larger number of patients

Sponsor also performed other analyses on the primary endpoint, for example, the responder's analysis. Significantly more patients had big improvement (>=4 unit improvement in OHSA 1) in droxidopa group compared with placebo group in the responder's analysis (Table 6).

Table 6. Responder's Analysis on OHSA 1 at Week 1

	OHSA Item 1 Unit Improvements		
	Placebo (N=78) N (%)	Droxidopa (N=69) N (%)	P-value
≥1 Unit Improvement from Baseline			0.118
Yes	46 (59.0)	50 (72.5)	
No	32 (41.0)	19 (27.5)	
≥2 Unit Improvement from Baseline			0.013
Yes	32 (41.0)	43 (62.3)	
No	46 (59.0)	26 (37.7)	
≥3 Unit Improvement from Baseline			0.027
Yes	23 (29.5)	33 (47.8)	
No	55 (70.5)	36 (52.2)	
≥4 Unit Improvement from Baseline			0.032
Yes	18 (23.1)	28 (40.6)	
No	60 (76.9)	41 (59.4)	

[Source: Sponsor's Clinical Study Report Table 11-6, verified by the reviewer]

The primary analysis used Full Analysis Set, which consisted of patients who took at least one dose of study drug and had OHSA Item 1 score at Week 1. 20 patients randomized to droxidopa were excluded from the primary analysis and only 7 patients in placebo were excluded. Droxidopa group had more dropouts during the titration phase. Table 7 listed the dropout reasons for these patients. Among treated patients, 6 placebo patients and 18 droxidopa patients discontinued study before Week 1.

Table 7. Discontinuation Reason for Patients Excluded from Full Analysis Set

Discontinuation Reason	Placebo	Droxidopa
Not treated	1	2
Treatment Failure	0	1
Adverse Event	4	6
Lack of Efficacy	0	3
Protocol Violation	0	1
Patient Withdrew Consent	0	3
Investigator Decision	0	2
Other	2	2
Total	7	20

The sponsor argued that the patient discontinuation rate was inflated in 306B. The interim analysis for 306 only included the patients who completed titration phase and finished the study. So patients who enrolled early but discontinued the study prior to completing titration were not included in the interim analysis. A total of 8 patients enrolled in the trial and dropped out during titration phase prior to the interim cut-off date. They were excluded from interim analysis and therefore were included in Study 306B. 7 out of the 8 patients were in droxidopa group. But even by excluding these 8 patients, Study 306B still had 5 placebo patients and 11 droxidopa patients who discontinued study prior to Week 1. The imbalance remained. In addition, one of the five placebo patients was treated with droxidopa although the planned treatment was placebo. So 4 patients treated with placebo and 12 patients treated with droxidopa discontinued study prior to Week 1. In fact, both patients enrolled earlier and patients enrolled later in the study showed similar pattern that droxidopa group had more dropouts.

Since OHSA Item 1 score was not measured during titration phase, patients with missing OHSA Item 1 at Week 1 only had baseline OHSA Item 1 score. One simple way to impute the missing data was to carry forward the baseline observations. The treatment effect was -0.45 with 95% confidence interval (-1.2, 0.3). This is not surprising since droxidopa group had more missing data and the imputation would bring more zeros to the droxidopa group.

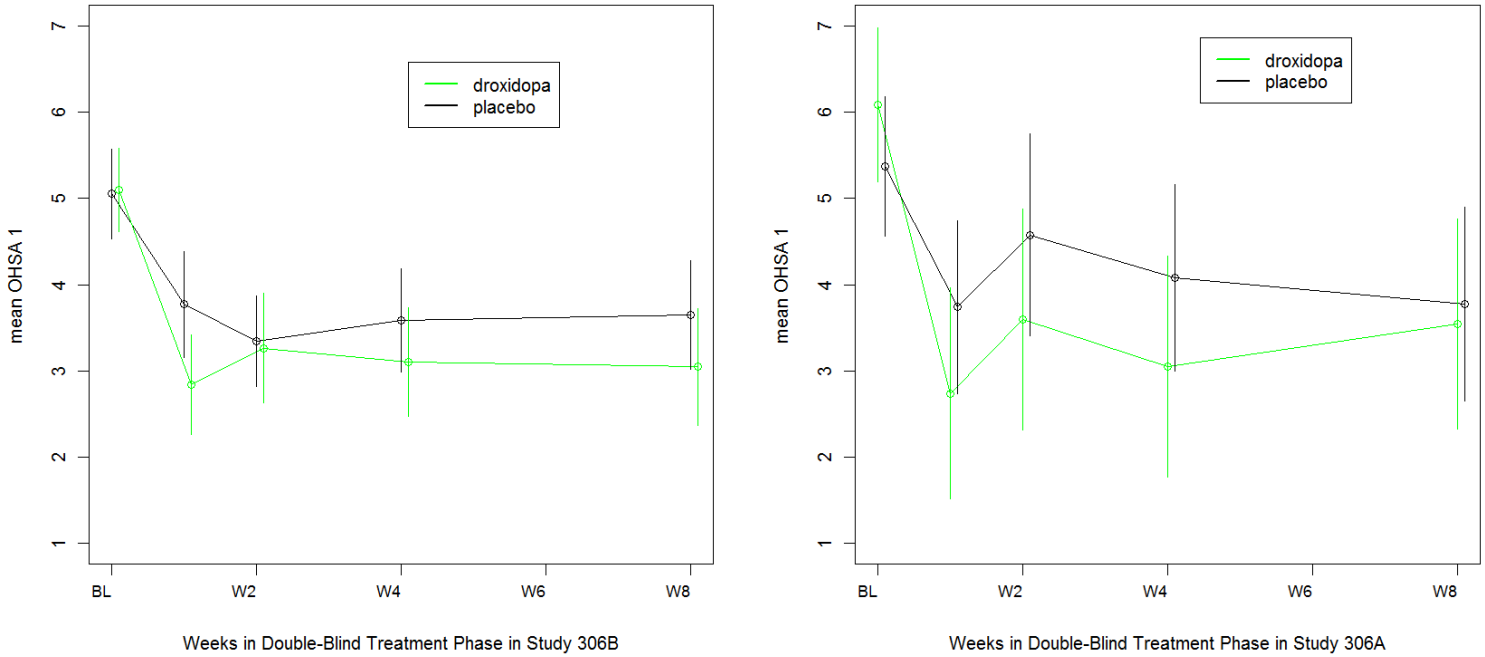
The reviewer examined the durability of the treatment effect on OHSA Item 1 by looking at its change from Baseline to Weeks 2, 4 and 8 (Table 8 and Figure 4). The treatment effect on OHSA Item 1 almost completely diminished at Week 2 and the treatment effect in Week 4 and Week 8 were also less than in Week 1.

Figure 5 and **Figure 6** displayed OHQ composite score and standing SBP (lowest between 0 and 3 minutes of standing in 306B, 3 minutes of standing in 306A) by visit. Study 306A showed almost no effect in OHQ composite score, which was the reason that DMC recommended terminating the trial for futility in 2011. Depending on the visits, the treatment effect in change of OHQ composite score varied between 0.4 to 0.7 unit in Study 306B (Table 12). The standing systolic blood pressure (lowest between 0 and +3 minutes of standing) had 5.4 mmHg more increase in change from Baseline to Week 1 in droxidopa group when compared with placebo. The treatment effect, however, did not seem to sustain through the 8-week treatment period for both Study 306A and Study 306B (**Table 13**). It is questionable whether droxidopa has any long-term clinical benefits.

Table 8. Summary on OHSA Item 1 Score at Weeks 2, 4 and 8

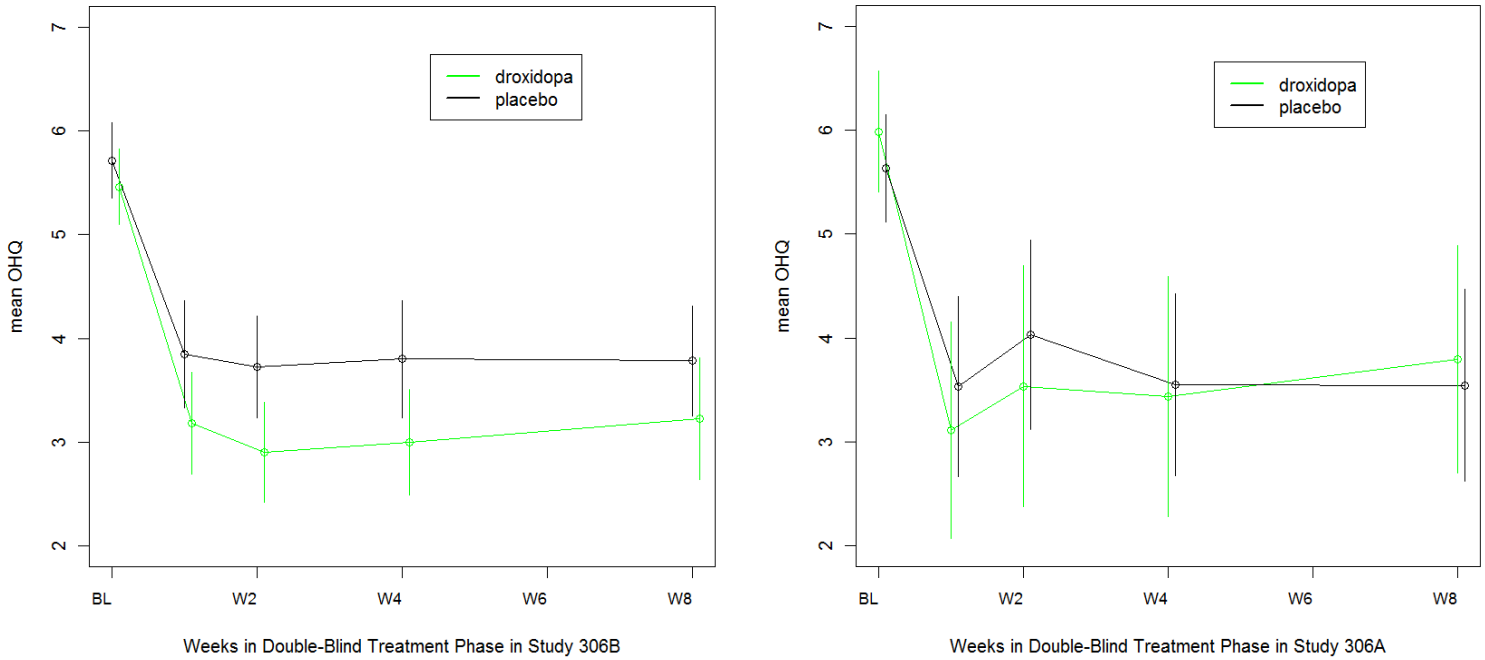
	Droxidopa			Placebo		
	N	Mean	STD	N	Mean	STD
Week 2	68	3.3	2.69	75	3.3	2.32
Change from Baseline to Week 2	68	-1.9	2.86	75	-1.6	2.97
Least square mean difference	-0.12 with 95% CI (-0.93, 0.69)					
p-value from ANCOVA	0.77					
Week 4	67	3.1	2.64	73	3.6	2.6
Change from Baseline to Week 4	67	-2	3.08	73	-1.5	2.74
Least square mean difference	-0.5 with 95% CI (-1.33, 0.36)					
p-value from ANCOVA	0.26					
Week 8	63	3	2.75	68	3.6	2.64
Change from Baseline to Week 8	63	-2.1	3.03	68	-1.5	2.91
Least square mean difference	-0.6 with 95% CI (-1.49, 0.30)					
p-value from ANCOVA	0.19					

Figure 4. OHSA 1 by Visit



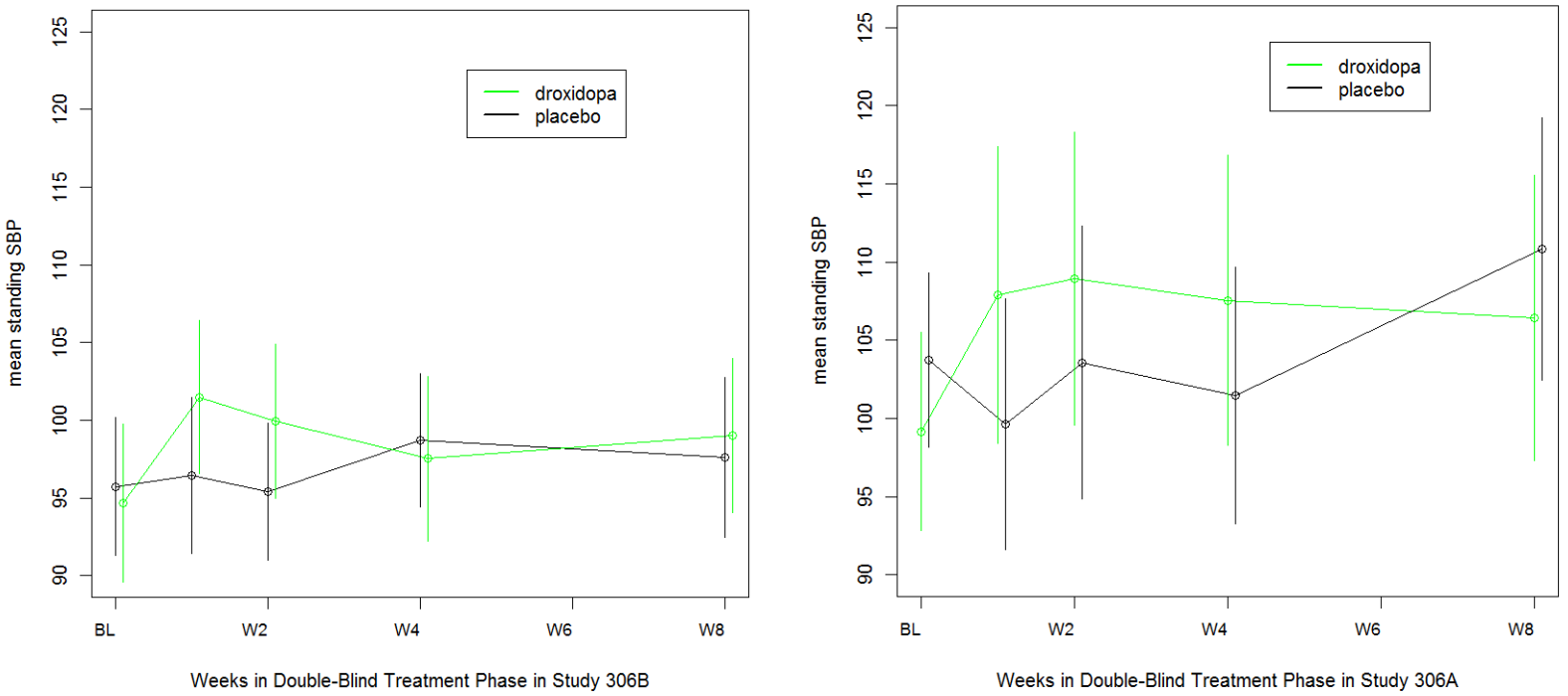
* left side is the mean OHSA Item 1 score by visit in each treatment group in Study 306B, right side is the mean OHSA Item 1 score by visit in each treatment group in Study 306A. Vertical lines are the 95% confidence interval of the mean OHSA Item 1 score for each individual treatment group at each specific visit

Figure 5. OHQ by Week



* left side is the mean OHQ composite score by visit in each treatment group in Study 306B, right side is the mean OHQ composite score by visit in each treatment group in Study 306A. Vertical lines are the 95% confidence interval of the mean OHQ composite score for each individual treatment group at each specific visit

Figure 6. Standing SBP by Week



* left side is the mean of the lowest standing SBP between 0 and +3 minutes of standing by visit in each treatment group in Study 306B, right side is the mean of standing SBP at 3 minutes of standing by visit in each treatment group in Study 306A. Vertical lines are the 95% confidence interval of the mean standing SBP for each individual treatment group at each specific visit

Patient-reported fall was once the primary endpoint. The sponsor showed that in Study 306B droxidopa patients experienced a lower total number of falls during the treatment period when compared with placebo patients (**Table 9**). By further examining the data, the reviewer noticed that patient 122013 and patient 146007 in placebo group had 118 and 358 reported falls, respectively. If excluding the two patients, the total number of falls in placebo group reduced to 240 compared with 229 reported falls in droxidopa group. The treatment difference in the total number of falls disappeared.

Table 9. Summary on Patient-Reported Falls

Analysis	Placebo (N=78)[2]	Droxidopa (N=69)[1]
Total Number of Falls, n	716	229
Percentage of Patients with ≥1 Fall[3], n (%)	47 (60.3)	40 (58.0)
Mean Patient Rate of Falls Per Patient-Week[4]	2.0 (12.95)	0.4 (0.84)

[Source: Sponsor's clinical study report Table 11-11, verified by the reviewer]

Study 306 went through a number of major changes during its course of conduct including changing the primary endpoint twice, splitting into Study 306A and Study 306B, changing sample size, and discovering inappropriate access to the treatment code for all study patients. **Table 10** summarized the chronicle of Study 306. The division had concerns over a number of major changes on the study design, especially towards the end of the study, which would undermine the creditability of the study results. Although the sponsor provided documents on their blinding process, it was impossible to be aware of every non-electronic communication occurred.

The sponsor also performed a post-interim sensitivity analysis on efficacy endpoints that included 121 patients to show that the post-interim results were consistent with the whole study. Based on the order of enrollment date, the Post-interim Analysis Set would include all FAS patients who were randomized after November 10, 2010. The cutoff date for the interim analysis, however, was December 14, 2010. Since maintaining treatment blinding was the concern, every patient who was randomized before the conduct of interim analysis should be excluded for sensitivity analysis.

So reviewer performed a similar post-interim analysis by including only patients who were randomized after December 14, 2010. A total of 113 patients were included in the reviewer's analysis. The results were similar to the sponsor's results on 121 patients and were consistent with the whole population (**Table 11**). The reviewer also performed similar subset analysis at different time points to further examine the data consistency. The treatment effects in various measurements were all trending in the right direction but the magnitude of the treatment effect tended to be less for the patients who enrolled later during the trial. For example, the estimate on the change in OHSA Item 1 score from Baseline to Week 1 was 0.6 by excluding all patients who were randomized before the inappropriate access to treatment code was revoked.

An interesting finding was that the treatment effect in the patient-reported CGI-I always was less than in the clinician-reported CGI-I (**Table 11**), which may be an indication of bias on one of the measurements.

Table 10. Timeline on Major Events

Date	Event
March 10, 2010	Study 306 protocol version 1: total sample size was at least 84. Primary endpoint was OHQ composite score at Week 8. The study was multi-national and it had no interim analysis
September 1, 2010	Study 306 protocol version 2: The study was changed to US only
November 19, 2010	Study 306 protocol version 3: Interim analysis at 60% information time (N=50) was added to re-assess treatment effect. This may result in sample size increase up to a maximum of 192
December 14, 2010	Cut-off date for 306 interim analysis. 94 patients were enrolled. The analysis included the first 51 patients who completed End of Study visit. PPD extraction Team extracted data from 92 patients into a Blinded Project Area where the unblinded DMC team have access
January 25, 2011	DMC met and recommended to stop Study 306 due to futility. 113 patients were enrolled into the study
February 9, 2011	PPD informed Chelsea that the unblinded statistical team may have been provided with access to the randomization codes for all Study 306 subjects.
February 23, 2011	Enrollment resumed for Study 306
March 2, 2011	PPD confirmed that unblinded statistical team did have access to the treatment code for all 306 subjects. The access was revoked. 118 patients were enrolled in the study by now.
April 11, 2011	FDA advised on protocol amendment submitted on March 16, 2011 that "Study NOH306B will not be accepted by the Division as supportive of efficacy"
May 12, 2011	Study 306 protocol version 4: The primary endpoint was changed to difference in patient reported falls at Week 8. The study was split into Study 306A (N=51) and Study 306B (N=160). No interim analysis was planned for Study 306B.
September 28, 2011	Chelsea submitted NDA including Study 301 and Study 302
November 5, 2011	Study 306 protocol version 5: The primary endpoint was changed to OHSA Item 1 at Week 1 for Study 306B and sample size was increased to 200
March 28, 2012	Complete response letter was issued
May 10, 2012	159 patients enrolled in Study 306B
May 31, 2012	Chelsea proposed to use Study 306B to fulfill FDA's requirement for additional confirmatory trial

June 29, 2012	FDA expressed concern on Study 306B, stating that it is impossible to know with certainty that interim results did not influence decisions to change the primary endpoint of Study 306B
August 10, 2012	Last patient enrolled in Study 306B. The sponsor announced to stop patient enrollment in July 2012.

Table 11. Comparison of Efficacy Results at Different Time Point

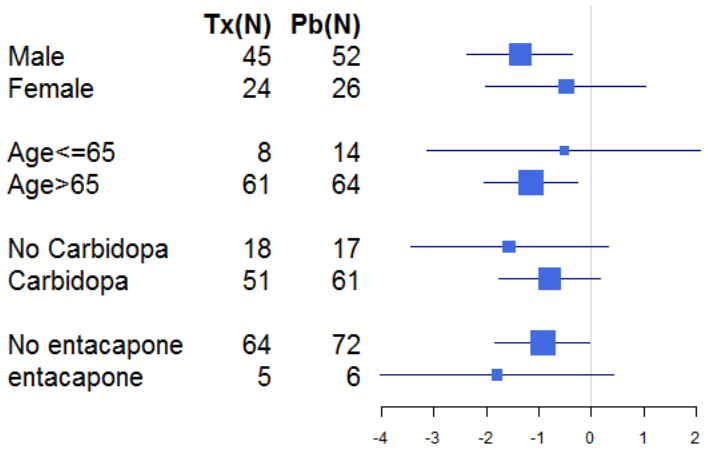
	Whole Study Population		Sponsor's Post Interim Analysis		Reviewer's Post Interim Analysis		Revoking Access to Treatment Code		Changing Primary Endpoint	
			After Nov 10, 2010		After Dec 14, 2010		After March 2, 2011		After May 12, 2011	
	N=147		N=121		N=113		N=93		N=71	
	trt eff est	CI	trt eff est	CI	trt eff est	CI	trt eff est	CI	trt eff est	CI
OHSA Item 1: Mean change from baseline at Week 1	-0.9	(-1.8, 0.1)	-1.1	(-2.0, -0.1)	-1.0	(-2.0, -0.05)	-0.6	(-1.7, 0.5)	-0.7	(-2.0, 0.6)
Lowest standing SBP between 0 to 3 minutes at Week 1	5.4	(-0.5, 11.3)	5.8	(-0.9, 12.4)	5.0	(-2.0, 12.0)	2.5	(-5.0, 10.0)	0.8	(-8.5, 10.1)
OHQ mean change from baseline at Week 1	-0.6	(-1.2, 0.1)	-0.7	(-1.5, 0.03)	-0.7	(-1.4, 0.1)	-0.4	(-1.2, 0.4)	-0.3	(-1.3, 0.7)
Clinician-reported CGI-S at Week 1	-0.4	(-0.8, -0.05)	-0.5	(-0.9, -0.1)	-0.5	(-0.9, -0.1)	-0.4	(-0.9, 0.03)	-0.2	(-0.7, 0.3)
Patient-reported CGI-S at Week 1	-0.4	(-0.8, 0.02)	-0.5	(-0.9, -0.04)	-0.5	(-0.9, -0.02)	-0.4	(-1.0, 0.1)	-0.2	(-0.8, 0.4)
Clinician-reported CGI-I at Week 1	-0.5	(-0.9, -0.1)	-0.6	(-1.0, -0.2)	-0.7	(-1.1, -0.2)	-0.5	(-1.0, -0.1)	-0.4	(-1.0, 0.1)
Patient-reported CGI-I at Week 1	-0.2	(-0.5, 0.1)	-0.3	(-0.7, 0.01)	-0.3	(-0.7, 0.02)	-0.2	(-0.6, 0.2)	-0.2	(-0.7, 0.3)

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The population for Study 306B was predominantly white and all patients were enrolled in US. Therefore, no subgroup analyses on race and country were performed. Figure 7 showed results of some subgroup analyses.

Figure 7. Forest Plot on Subgroup Analyses



4.2 Other Special/Subgroup Populations

The reviewer specifically examined patients by whether they took entacapone or not and whether they took carbidopa/levodopa (Sinemet) since carbidopa and entacapone may modify the metabolism of droxidopa. The results were showed in the forest plot (Figure 7).

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

After changing the primary endpoint twice, the final primary efficacy endpoint was the mean change in the OHSA Item 1 from Baseline to Week 1. The sponsor concluded that the assumptions for the ANCOVA were not met and used non-parametric methodology for the primary analysis. Based on the sponsor's analysis, the droxidopa group had a treatment effect of

-1 with 95% confidence interval (-2.0, 0) when compared to placebo group in change of OHSA Item 1 score from Baseline to Week 1 and the p-value was 0.018. The reviewer, however, did not find any obvious deviation from ANCOVA assumptions. The treatment effect based on ANCOVA model was -0.94 and the p-value was 0.028. Both results were statistically significant.

Although statistically significant, the treatment effect on OHSA Item 1 at Week 1 seemed small at the presence of intra-subject variability, which was 2.9 based on reviewer's calculation.

The treatment effect at later weeks in the study was not so consistent. The treatment effect on OHSA Item 1 almost completely diminished at Week 2 and was also less at Week 4 and Week 8. The treatment effect in standing SBP did not sustain through the 8-week treatment period. This made it questionable whether droxidopa has any long term treatment effect.

Study 306 went through a number of major changes during its course of conduct including changing the primary endpoint twice, splitting into Study 306A and Study 306B, and changing the total sample size. In addition, it was discovered that the unblinded statistical team had access to the treatment codes for all Study 306 subjects rather than the 51 patients for the interim analysis. Although the access was later revoked, a considerable number of patients in Study 306 were already enrolled. In order to address the concerns on study conduct, the sponsor performed a post-interim sensitivity analysis to show that the study results remained consistent. The reviewer also performed similar analyses at additional time points, such as after revoking the access to treatment code and after changing to the final primary endpoint. The treatment effects in various measurements were all trending in the right direction but the magnitude of the treatment effect tended to be less for the patients who enrolled later during the trial.

Droxidopa group had more dropouts during the titration phase. 20 droxidopa patients were excluded from the primary analysis compared with only 7 placebo patients. Except for three untreated patients, the rest of these patients had missing OHSA Item 1 score at Week 1. Even if excluding 8 patients who enrolled earlier before the interim analysis, Study 306B still had 4 patients treated with placebo and 12 patients treated with droxidopa discontinued study prior to Week 1. The imbalance remained. It is concerning to see such imbalance of dropouts between treatment groups, especially if the data were not missing at random. The treatment effect of OHSA Item 1 became -0.45 with 95% confidence interval (-1.2, 0.3) if imputing missing data by carrying forward baseline observation (BOCF).

5.2 Conclusions and Recommendations

The droxidopa group had a statistically significant treatment effect over placebo group in the mean change in the OHSA Item 1 score from Baseline to Week 1. Other measurements at Week 1 were all trending in the right direction, though might not reach statistical significance.

However, the treatment effect on OHSA Item 1 at Week 1 seemed small when compared with intra-subject variability. It is also concerning to observe an imbalance of dropouts between treatment groups. The treatment effect of droxidopa did not seem to sustain through the 8-week

treatment period. This made it questionable whether droxidopa has any long term treatment effect.

The credibility of the study was also undermined by a number of major changes on the study design and the discovery of inappropriate access to the treatment codes of all study patients enrolled until March 2011. Sensitivity analyses were performed to include only patients enrolled after certain time point to examine the consistency of the study results. The treatment effects in various measurements were all trending in the right direction but the magnitude of the treatment effect tended to be less for the patients who enrolled later during the trial.

Overall, Study 306B alone did not seem to provide strong and robust evidence to support the efficacy of droxidopa in treating NOH, especially for long-term treatment.

APPENDIX

Table 12. Summary on OHQ Composite Score by Visit

	Droxidopa			Placebo		
	N	Mean	STD	N	Mean	STD
Baseline	69	5.5	1.54	78	5.7	1.64
Week 1	69	3.2	2.07	78	3.9	2.33
Change from Baseline to Week 1	69	-2.3	2.12	78	-1.9	2.39
Least square mean difference	-0.55 with 95% CI (-1.24, 0.14)					
p-value from ANCOVA	0.115					
Week 2	68	2.9	2.03	75	3.7	2.17
Change from Baseline to Week 2	68	-2.5	1.98	75	-2	2.26
Least square mean difference	-0.71 with 95% CI (-1.37, -0.06)					
p-value from ANCOVA	0.032					
Week 4	67	3	2.12	73	3.8	2.46
Change from Baseline to Week 4	67	-2.5	1.93	73	-1.9	2.28
Least square mean difference	-0.64 with 95% CI (-1.33, 0.05)					
p-value from ANCOVA	0.068					
Week 8	63	3.2	2.38	68	3.8	2.23
Change from Baseline to Week 8	63	-2.2	2.29	68	-2	2.18
Least square mean difference	-0.40 with 95% CI (-1.14, 0.38)					
p-value from ANCOVA	0.29					

Table 13. Summary on Standing SBP by Visit

	Droxidopa			Placebo		
	N	Mean	STD	N	Mean	STD
Baseline	69	94.7	21.5	78	95.7	20.1
Week 1	68	101.5	20.8	78	96.4	22.7
Change from Baseline to Week 1	68	6.4	18.9	78	0.7	20.2
Least square mean difference	5.4 with 95% CI (-0.5, 11.3)					
p-value from ANCOVA	0.07					
Week 2	68	99.9	20.9	75	95.4	19.6
Change from Baseline to Week 2	68	5.5	19.3	75	-0.6	20.3
Least square mean difference	5.4 with 95% CI (-0.3, 11.0)					
p-value from ANCOVA	0.06					
Week 4	65	97.5	21.9	73	98.7	18.7
Change from Baseline to Week 4	65	2.8	20.2	73	3	19.4
Least square mean difference	-0.7 with 95% CI (-6.4, 5.1)					
p-value from ANCOVA	0.82					
Week 8	64	99	20.3	69	97.6	21.8
Change from Baseline to Week 8	64	5	18.5	69	0.9	18.4
Least square mean difference	3.0 with 95% CI (-2.7, 8.8)					
p-value from ANCOVA	0.29					

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

/s/

JIALU ZHANG
12/03/2013

HSIEN MING J HUNG
12/03/2013

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

ADDENDUM

NDA/BLA Serial Number:	NDA 203-202 / SN 0000
Drug Name:	Droxidopa
Indication(s):	treatment of symptomatic neurogenic orthostatic hypotension (NOH) in patients with primary autonomic failure (Parkinson's Disease [PD], Multiple System Atrophy [MSA], and Pure Autonomic Failure [PAF]), Dopamine Beta-Hydroxylase (DBH) Deficiency, and Non-Diabetic Autonomic Neuropathy (NDAN)
Applicant:	Chelsea Therapeutics, Inc
Date(s):	Date of Document: September 28, 2011 PDUFA due date: March 28, 2011
Review Priority:	Priority
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	James Hung, Ph.D.
Medical Division:	Division of Cardiovascular and Renal Products, HFD-110
Clinical Team:	Melanie Blank, M.D.
Project Manager:	Anna Park, Pharm.D.

Keywords:
Responder analysis, effect size

Chelsea is seeking an indication on droxidopa in treating symptomatic neurogenic orthostatic hypotension (NOH) in patients with primary autonomic failure (Parkinson's Disease, Multiple System Atrophy) and pure autonomic failure (Dopamine Beta Hydroxylase Deficiency and Non-Diabetic Autonomic Neuropathy). The primary statistical review was completed on January 23, 2012. This addendum is to summarize some additional analyses performed by the reviewer.

1. Graphic display on reduction of the scores

Figure 1 shows the change of each component of the OHQ score as well as the OHQ composite score from Randomization Visit to End of Study (EOS) Visit. The height of the stacked bars is the score at Randomization Visit and the green/light green portion is mean decrease in each treatment group. The asterisks mark the individual items that have p-value < 0.05 in change from Randomization Visit in two treatment groups by ANCOVA model. The histogram provides a visual display on the amount of reduction relative to the baseline (Randomization visit). Similar histograms were also plotted for Study 302 and Study 303 (Figure 2 and Figure 3).

Figure 1 Score Change from Randomization Visit to EOS Visit in Study 301

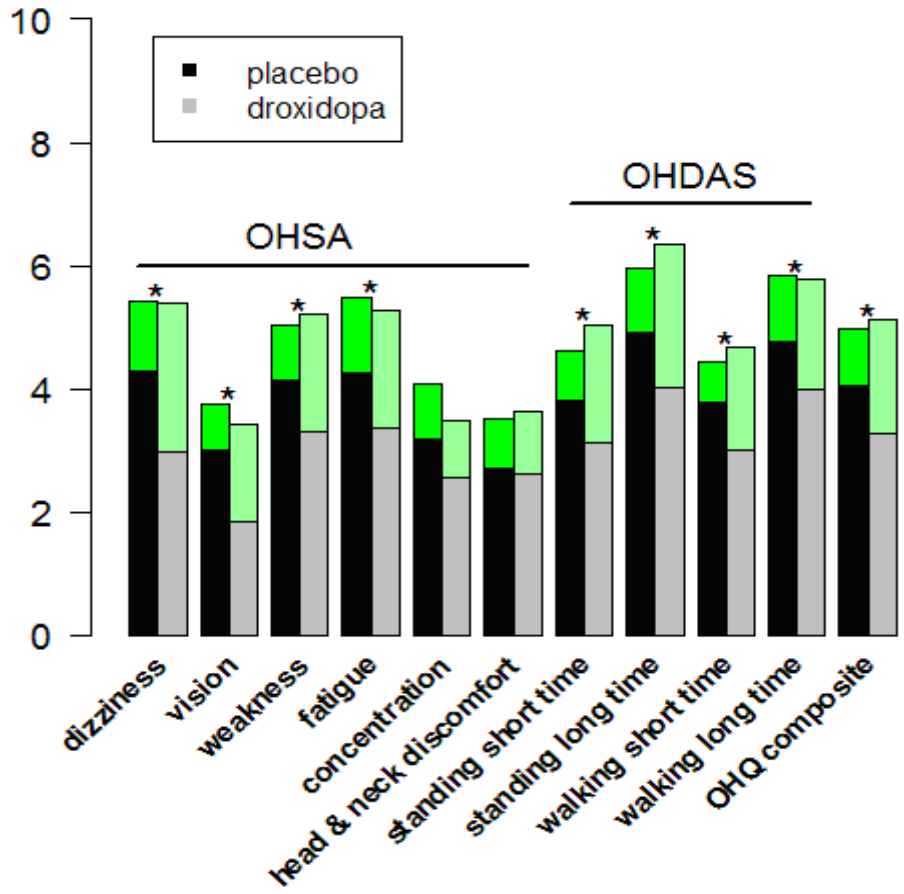
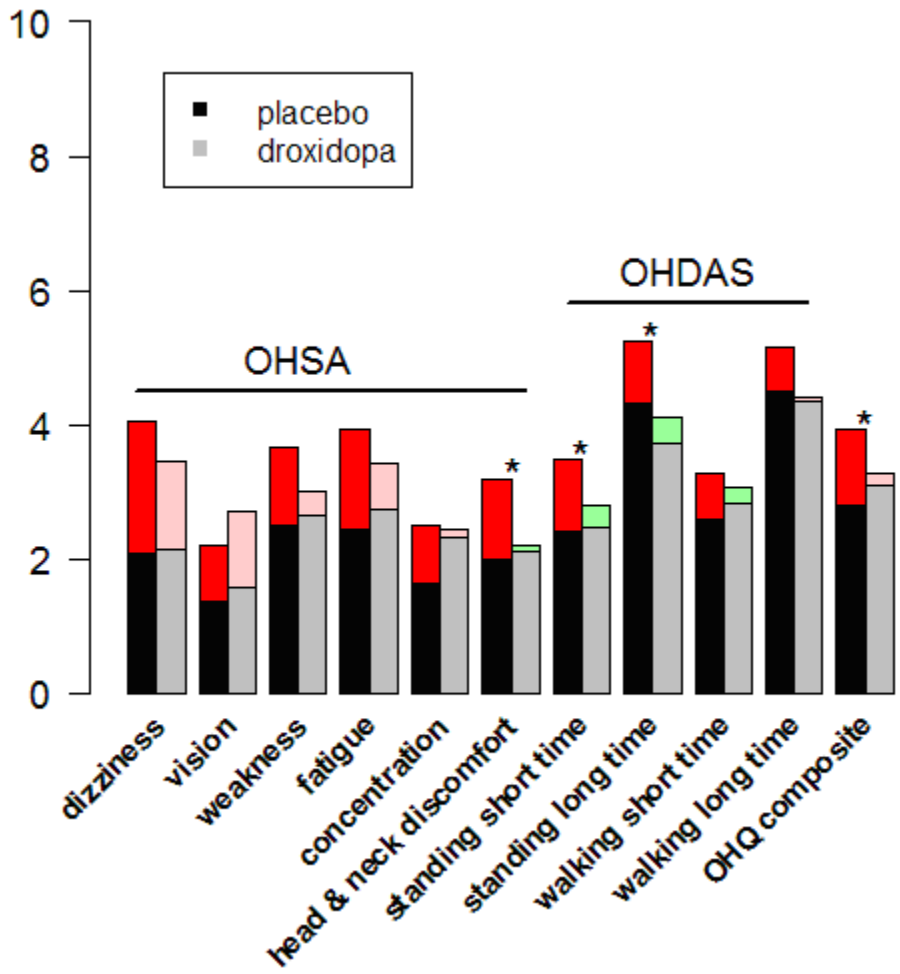
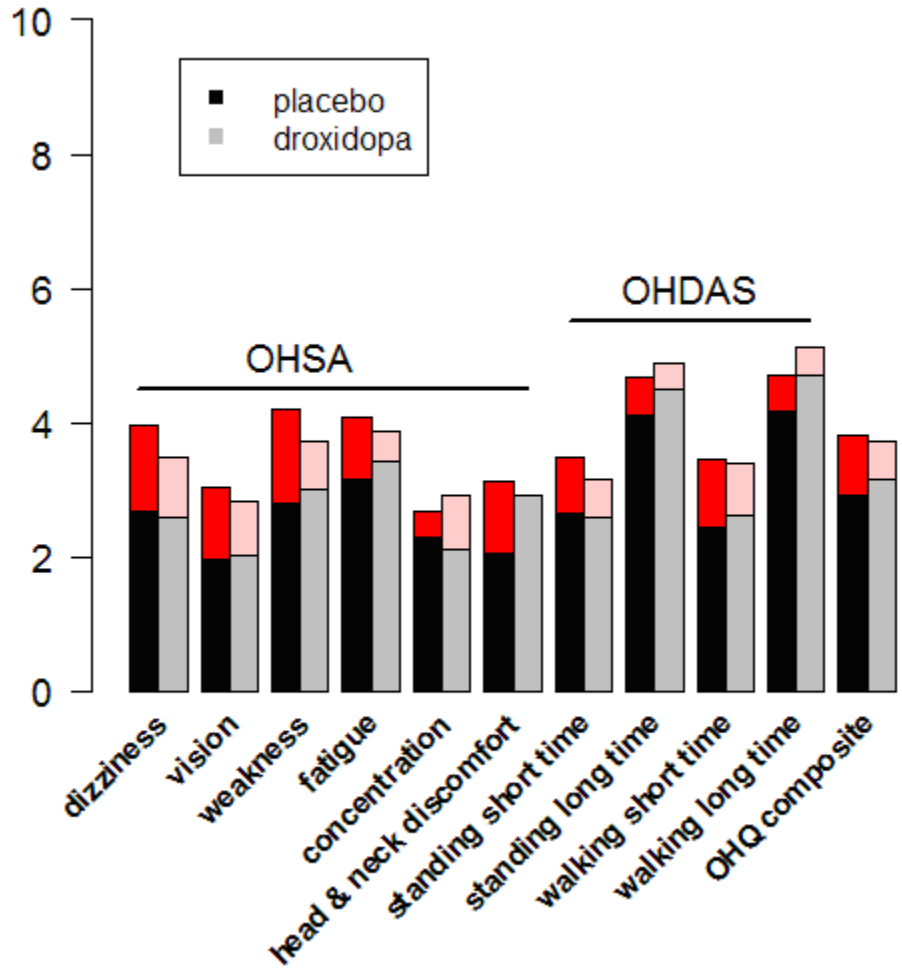


Figure 2 Score Change from Randomization Visit to EOS Visit in Study 302



1. Asterisks indicate p-value<0.05 in the changes of the score in two treatment groups by ANCOVA model. Red / pink bars shows mean increase of a score from Randomization visit to EOS visit. Green / light green bars show the mean decrease of a score from Randomization visit to EOS visit.

Figure 3 Score Change from Randomization Visit to EOS Visit in Study 303



1. Red / pink bars show the mean increase of a score from Randomization visit to EOS visit.

2. OHQ / OHSA Item 1 and other measurements

The reviewer performed additional analyses to better understand how the mean OHQ change and mean OHSA item 1 change associate with other measurements.

2.1 Change of SBP and change of OHQ / OHSA Item 1

The change of SBP from randomization to EOS and change of OHQ from randomization to EOS were plotted in Figure 4. The black line on the graph is from the linear regression model. R square from the linear model is 0.09 and the slope estimate is -3.1. Overall, the

relationship between SBP change and OHQ change is weak. Change in OHQ score only explains about 9% of the variation in change of SBP in the linear model. The diagnostic plots indicate that several linear model assumptions are not valid, such as, deviation from normal distribution in the residuals (heavy tails) and non-constant variance. So the linear model may not be a good fit. The green curve in the graph is based on LOWESS smoothing. Surprisingly, the LOWESS smoothing curve does not deviate much from the linear regression line. Figure 5 is similar but this one is to examine the relationship between SBP and OHSA item 1. The linear model has R square of 0.11 and the slope estimate is -2.2. The change in OHSA item 1 appears to have a wider range than the change in OHQ score. The relationship between SBP change and OHSA item 1 change is also weak.

Figure 4 Distribution on Change of OHQ Score in Placebo and Droxidopa in Study 301

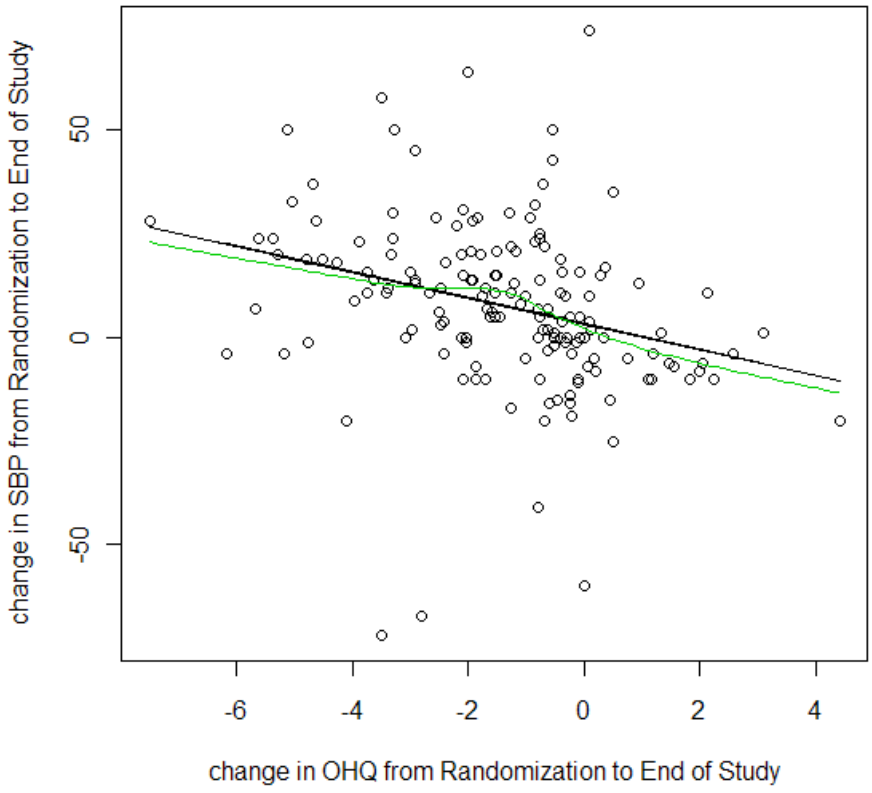
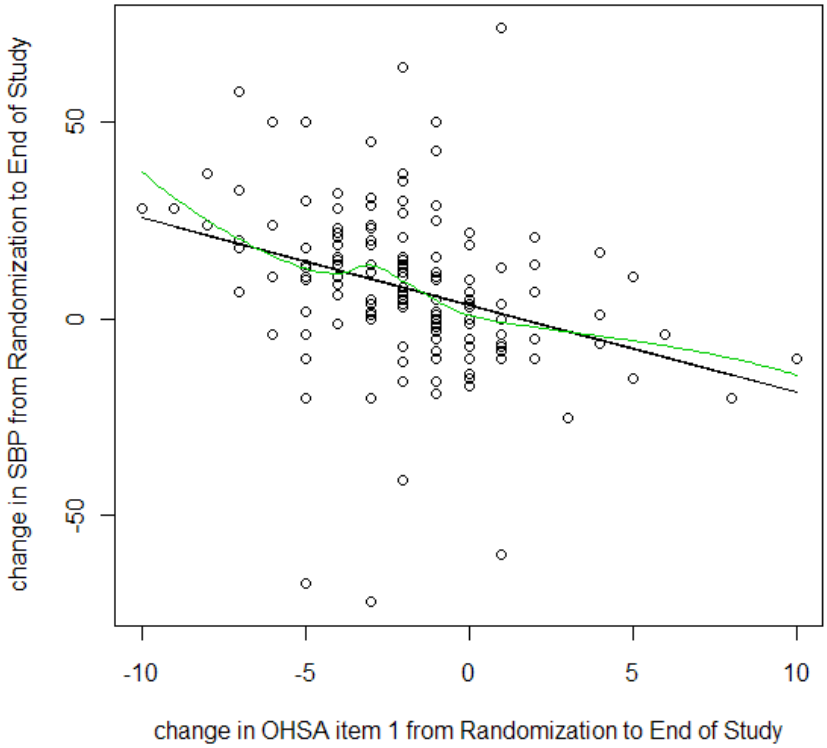


Figure 5 Distribution on Change of OHSA 1 Score in Placebo and Droxidopa in Study 301



2.2 CGI-I / CGI-S and change of OHQ / OHSA Item 1

CGI-I is a 7-point scale ranging from a score of 1 (very much improved) to 7 (very much worse), with no change in the middle, and assesses the improvement in relation to the baseline evaluation. CGI-I was only measured on Randomization Visit (visit 4) and EOS visit (visit 5). The reviewer focused on the mean changes in OHQ/OHSA 1 scores from baseline to EOS visit since patients went through a double-blind, randomized treatment period before the EOS visit and results can be less subjective. It is noticed that most differences of mean OHQ changes (Table 1) or mean OHSA item #1 changes (Table 2) between two consecutive CGI-I cagetories are above 1. Except between categories “no change” and “minimally worse”, the difference of OHQ scores between two consecutive CGI-I categories ranges from 0.94 to 1.95. For example, according to Table 2, in order to move from “no change” to “minimally improved”, the mean change in OHQ scores need to increase 1.69. Similarly, the difference of mean OHSA item #1 change two consecutive CGI-I categories ranges from 1.49 to 2.25.

The CGI-I measurement, on the other hand, can be confusing to patients since the questionnaire asked subjects to compare with their conditions at baseline visit (Visit 2), not the previous visit. This may reduce the reliability of the results.

Table 1 Mean OHQ change scores by CGI-I patient reported changes in study 301

	CGI-I	N	Mean OHQ Change	SD	95% CI
Visit 5	Very much improved	28	-3.91	2.05	(-4.70, -3.11)
	Much improved	36	-2.87	1.42	(-3.35, -2.39)
	Minimally improved	48	-1.93	1.52	(-2.38, -1.49)
	No change	30	-0.24	1.15	(-0.67, 0.19)
	Minimally worse	12	-0.41	1.82	(-1.57, 0.74)
	Much worse	5	1.54	1.70	(-0.56, 3.65)
	Very much worse	-	-	-	-

[Source: Table 1.3.1 from sponsor response to FDA request submitted on 1/9/2012, verified by the reviewer]

Table 2 Mean OHSA Item #1 change scores by CGI-I patient reported changes in study 301

	CGI-I	N	Mean OHSA Item #1 Change (SD)	95% CI	Effect Size
Visit 5	Very much improved	28	-5.68(2.44)	(-6.62, -4.73)	-2.667
	Much improved	36	-3.97(1.84)	(-4.60, -3.35)	-1.714
	Minimally improved	48	-2.48(1.68)	(-2.97, -1.99)	-1.284
	No change	30	-0.23(2.14)	(-1.03, 0.57)	-0.090
	Minimally worse	12	-0.58(2.84)	(-2.39, 1.22)	-0.302
	Much worse	5	1.00(1.87)	(-1.32, 3.32)	0.461
	Very much worse		-	-	-

[Source: Table 1.1.1 from sponsor response to FDA request submitted on 1/31/2012, verified by the reviewer]

CGI-S is a 7-point scale ranging from a score of 1 (no symptoms) to 7 (severe symptoms). A reduction in score over a period of time would be considered an improvement in symptoms. Table 3 shows the mean OHQ change corresponding to the CGI-S change between Randomization Visit and EOS visit. The mean OHQ change corresponds to no change in CGI-S between Randomization Visit and EOS Visit has a 95% CI excluding 0, which does not seem sensible. This illustrates how variable these measurements can be.

Table 3 Mean OHQ change scores by changes in patient reported CGI-S in study 301

	CGI-S	N	Mean OHQ Change	SD	95% CI
Visit5-V4	Improve>=2 grade	47	-2.68	1.73	(-3.18, -2.17)
	Improve 1 grade	37	-1.56	1.43	(-2.04, -1.08)
	No change	54	-0.87	1.44	(-1.26, -0.47)
	Worse 1 grade	19	0.57	1.36	(-0.09, 1.23)
	Worse>=2 grade	4	1.78	2.12	(-1.59, 5.15)

[Source: Table 2.3.1 from sponsor response to FDA request submitted on 1/9/2012, verified by the reviewer]

Ideally, the mean change of symptom scores between categories of global assessments should be consistent among trials. The reviewer further looked at Study 302. The mean OHQ / OHSA 1 change between categories indicating worsening of symptoms does not seem sensible (the mean change in OHQ in "very much worse" category is better than the mean change in OHQ in "much worse" category). This is probably due to the small number of subjects in each category. For the categories that contains more than 10 subjects, the difference of mean OHQ change between two consecutive categories ranges from 0.65 to 1.54 and the difference of mean OHSA 1 change ranges from 1.03 to 1.87. Theoretically, Study 302 and Study 301 should have very similar results. Comparing to the results in Study 301, the measurements again seem quite variable. Nevertheless, the information here can be used to design the next trial.

Table 4 Mean OHQ change scores by CGI-I patient reported changes in study 302

	CGI-I	N	Mean OHQ Change	SD	95% CI
Visit 5	Very much improved	18	-4.68	2.76	(-6.05, -3.31)
	Much improved	18	-3.44	1.76	(-4.31, -2.56)
	Minimally improved	33	-2.79	2.01	(-3.50, -2.07)
	No change	10	-1.25	1.73	(-2.49, -0.01)
	Minimally worse	9	-1.03	2.18	(-2.71, 0.64)
	Much worse	6	1.68	1.72	(-0.13, 3.49)
	Very much worse	5	1.01	1.91	(-1.36, 3.38)

[Source: Table 1.3.2 from sponsor response to FDA request submitted on 1/9/2012, verified by the reviewer]

Table 5 Mean OHSA Item # 1 change scores by CGI-I patient reported changes in study 302

	CGI-I	N	Mean OHSA Item #1 Change (SD)	95% CI
Visit 5	Very much improved	18	-5.56(3.01)	(-7.05, -4.06)
	Much improved	18	-4.00(2.40)	(-5.19, -2.81)
	Minimally improved	33	-2.97(2.54)	(-3.87, -2.07)
	No change	10	-1.1(2.23)	(-2.70, 0.50)
	Minimally worse	9	-1.22(2.44)	(-3.10, 0.65)
	Much worse	6	2.00(2.45)	(-0.57, 4.57)
	Very much worse	5	1.8(2.16)	(-0.89, 4.49)

[Source: Table 1.1.2 from sponsor response to FDA request submitted on 1/31/2012, verified by the reviewer]

Table 6 Mean OHQ change scores by CGI-S patient reported changes in study 302

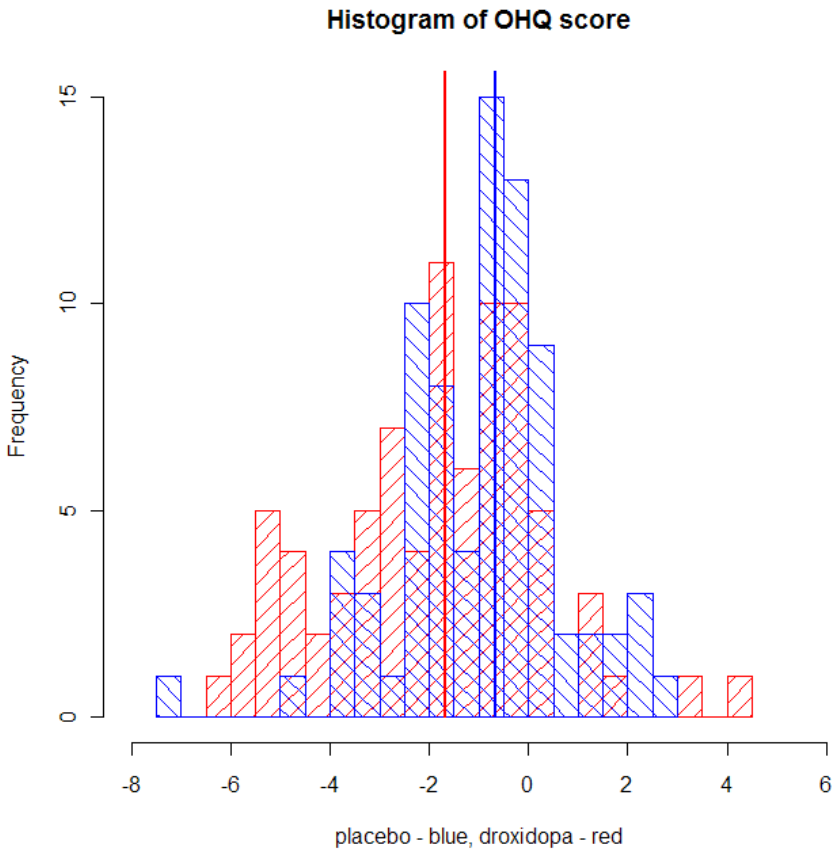
	CGI-S	N	Mean OHQ Change	SD	95% CI
Visit5-V4	Improve>=2 grade	12	-0.89	1.62	(-1.92, 0.14)
	Improve 1 grade	13	-0.57	1.72	(-1.61, 0.46)
	No change	36	0.14	1.73	(-0.44, 0.73)
	Worse 1 grade	13	1.54	2.17	(0.22, 2.85)
	Worse>=2 grade	25	2.56	2.47	(1.54, 3.58)

[Source: Table 2.3.2 from sponsor response to FDA request submitted on 1/9/2012, verified by the reviewer]

3. Distribution of OHQ / OHSA Item 1 scores and responder analysis in Study 301

Figure 6 shows a bar plot of change of OHQ score from Randomization Visit to End of Study Visit. Droxidopa group is shown in red and placebo in blue. Kolmogorov-Smirnov test has a p-value of 0.017 indicating that two are not from the same distribution. The two vertical lines are the medians of the change of OHQ score in two treatment groups. From the graph, most red bars on the left side of the red vertical line are higher than the blue bars. This should indicate a shift in the distribution since the numbers of subjects are quite similar in two groups (82 in droxidopa and 80 in placebo). The CDF plot also confirms that.

Figure 6 Distribution on Change of OHQ Score in Placebo and Droxidopa in Study 301



Similar pattern is observed in OHSA 1 score.

Figure 7 Distribution on Change of OHSA 1 Score in Placebo and Droxidopa in Study 301

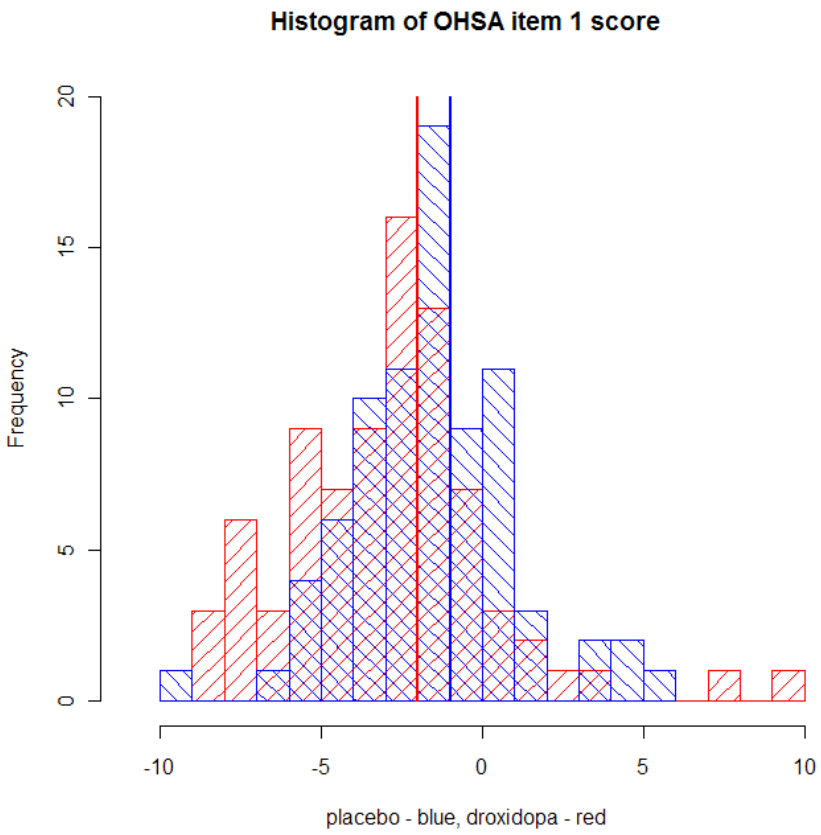
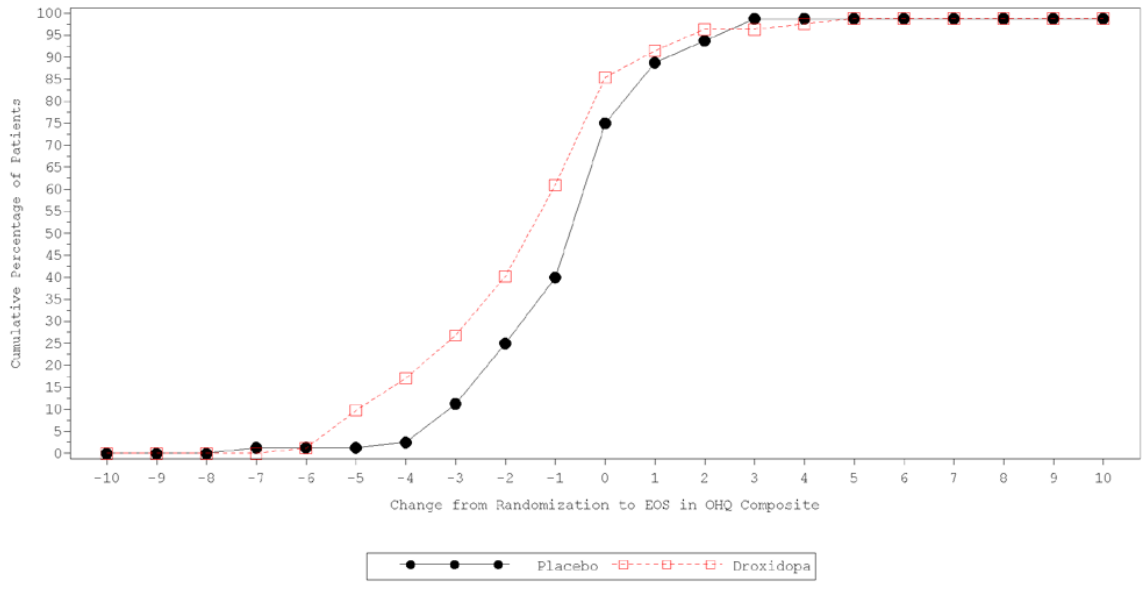


Figure 8 OHQ Cumulative Distribution on Placebo and Droxidopa in Study 301



[Source: Figure 4-1 in Sponsor's Clinical Overview, verified by the reviewer]

Responder analyses were performed by using different thresholds (Table 7). The p-values are based on Pearson's Chi-Square test. The droxidopa group has significantly more responders no matter what threshold to choose.

Table 7 Responder Analyses in Study 301

	droxidopa		placebo		
Responder definition	responder	nonresponder	responder	nonresponder	p-value
Reduction of OHQ >=1	51	31	33	47	0.0076
Reduction of OHQ >=2	34	48	21	59	0.04
Reduction of OHQ >=3	23	59	10	70	0.014
Reduction of OHQ >=4	15	67	3	77	0.003
Reduction of OHSA 1 >= 1	66	16	52	28	0.027
Reduction of OHSA 1 >= 2	53	29	33	47	0.003
Reduction of OHSA 1 >= 3	37	45	22	58	0.02
Reduction of OHSA 1 >= 4	28	54	12	68	0.005

4. Additional findings in Site 507

The Advisory Committee meeting was held on February 23, 2012. The committee members seemed to be impressed that some patients in the droxidopa group improved considerably (reduction of OHQ score can be larger than 4). After the Advisory Committee meeting, it was found that 6 out of the 15 subjects who had huge treatment effect in droxidopa group (reduction of OHQ score of more than 4) were from a single site from Ukraine. This Site 507 had 16 subjects enrolled and was the largest site for this trial. Further examination on that site illustrated strikingly homogeneous treatment effect in both the droxidopa and placebo groups (Table 8). It also seems very unusual that the site enrolled subjects in a short period given that this is a rare disease. The site was selected for inspection earlier but the inspector did not have significant finding that may impact data reliability.

Table 8 Subjects in Site 507

PTID	Screening Date	Date of first dose	Random Date	Treatment	OHQ			OHSA item 1				3 minute post standing SBP				CGI-I
					eos	rand	diff	baseline	RAND	EOS	diff	BL	RAND	EOS	diff	EOS
507018	8/6/2009	8/11/2009	8/21/2009	Droxidopa	1.8	5.1	-3.3	8	7	0	-7	96	98	118	20	1
507020	8/12/2009	8/19/2009	8/31/2009	Droxidopa	0.8	3.1	-2.4	6	7	0	-7	98	101	119	18	2
507003	4/15/2009	4/21/2009	5/8/2009	Droxidopa	2.1	6.8	-4.6	8	9	0	-9	92	93	121	28	1
507005	4/16/2009	4/22/2009	5/7/2009	Droxidopa	2.5	7.2	-4.7	9	8	0	-8	93	95	132	37	1
507007	5/13/2009	5/18/2009	6/1/2009	Droxidopa	1.9	7.0	-5.1	9	8	2	-6	103	98	148	50	1
507015	5/26/2009	6/2/2009	7/3/2009	Droxidopa	1.1	6.2	-5.0	7	8	1	-7	91	92	125	33	1
507016	6/3/2009	6/9/2009	6/24/2009	Droxidopa	1.6	7.0	-5.4	8	7	1	-6	90	96	120	24	2
507004	4/15/2009	4/21/2009	5/7/2009	Droxidopa	2.3	5.8	-3.5	8	8	1	-7	88	94	152	58	2
507009	5/14/2009	5/19/2009	6/2/2009	Droxidopa	1.3	6.5	-5.3	10	8	1	-7	91	94	114	20	1
507001	4/13/2009	4/16/2009	4/29/2009	Placebo	6.4	6.6	-0.2	9	8	9	1	90	93	89	-4	4
507002	4/13/2009	4/22/2009	5/7/2009	Placebo	6.3	8.4	-2.1	10	10	9	-1	90	92	92	0	4
507008	5/13/2009	5/20/2009	6/2/2009	Placebo	5.3	5.7	-0.4	7	8	6	-2	91	90	94	4	3
507010	5/20/2009	5/27/2009	6/9/2009	Placebo	5.4	5.7	-0.3	7	8	7	-1	87	89	89	0	4
507012	5/25/2009	6/1/2009	6/18/2009	Placebo	6.3	6.2	0.1	8	9	6	-3	89	91	93	2	3
507013	5/25/2009	5/28/2009	6/24/2009	Placebo	5.9	6.0	-0.1	8	9	7	-2	92	90	95	5	4
507014	5/25/2009	6/1/2009	6/30/2009	Placebo	4.6	7.0	-2.4	9	10	7	-3	96	95	99	4	3

Table 9 lists sites enrolled more than 3 subjects. The treatment effect of droxidopa is highly significant in Site 507 alone. If the 16 subjects from Site 507 were removed from ITT population and same analysis as the primary analysis was performed, the difference of mean change in OHQ score between the two treatment groups changed from 0.9 to 0.56 and the p-value changed from 0.003 to 0.082. The reviewer performed a simulation of 10,000 runs to randomly remove 16 subjects (9 from droxidopa and 7 from placebo). The probability of getting a p-value>0.05 is 0.0017 with maximum p-value of 0.759 from the simulation. So the probability of observing a p-value of 0.082 or greater by removing 16 subjects from the ITT population is less than 0.0001.

Table 9 Treatment effect by site

	Droxidopa			Placebo				
Site	N	Mean	STD	N	Mean	STD	Total N	p-value
507	9	-4.370	1.052	7	-0.763	1.038	16	<0.0001
505	6	-2.729	1.140	5	-1.067	1.512	11	0.007
125	3	-2.447	1.542	6	-1.424	1.803	9	0.505
607	3	-0.431	0.105	5	-0.133	0.555	8	0.771
100	2	-3.188	1.503	4	-1.917	3.738	6	0.774
103	3	-0.203	3.185	3	1.028	0.819	6	0.553
501	3	-0.722	1.672	3	-1.181	0.400	6	0.730
126	2	-0.958	1.237	3	-3.136	1.241	5	0.274
300	3	-2.458	2.394	2	-1.333	0.884	5	0.627
512	3	-1.806	0.646	2	-1.004	1.526	5	0.102
706	3	-0.722	0.907	2	-1.271	0.383	5	0.680
105	2	-0.917	1.296	3	-0.486	0.868	5	0.391
601	2	0.692	1.615	2	-0.417	3.771	4	0.498

In the primary statistical review, the reviewer concluded that only Study 301 demonstrated the efficacy of droxidopa. By then, the reviewer considered the results in Study 301 very much consistent and robust. With the new findings of Site 507, the robustness of results in Study 301 becomes questionable. If there is fraud involved in Site 507, then we no longer have a positive study. Even if the data of Site 507 are proved to be valid, it remains troublesome that the results of Study 301 are highly influenced by a single site in a foreign country, which may not have the same medical practice in US.

Therefore, the conclusion that we need additional study remains the same.

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

/s/

JIALU ZHANG
03/30/2012

HSIEN MING J HUNG
03/30/2012

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

CARCINOGENICITY STUDIES

NDA:	**203202**
Drug Name:	**Droxidopa (L-threo 3,4 Dihyroxyphenylse)**
Indications :	**Primary autonomic failure, dopamine failure, dopamine beta hydroxylase deficiency or non-diabetic neuropathy, and symptomatic neurogenic orthostatic hypotension**
Applicants:	**Chelsea Therapeutics, Inc.**
	Charlotte, North Carolina
	CRO: (b) (4)
Date(s):	**Initial Studies at CRO: 1987-1989**
	Re-analysis at CRO: 2011
	Current Submission: 28 September 2011
	To Reviewer: 13 October 2011
Review Priority:	**Accelerated (6 month)**
Biometrics Division:	**Division 6**
Statistical Reviewer:	**Steve Thomson**
Concurring Reviewers:	**Karl Lin, Ph.D.**
Medical Division:	**Cardiovascular and Renal Products**
Toxicologist Team:	**Donald Jensen, Ph.D.**
	Thomas Papoian, Ph.D.
Project Manager:	**Anna Park, Ph.D.**

Keywords: Carcinogenicity, Cox regresson, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY

Reports from a rat and a mouse study were provided. The original studies were conducted by (b) (4) at their research center in (b) (4) in 1987-1989 and 1988-1989, respectively. Reports for both studies were completed in 1991. The orginal rat report states that the object of this study "was to assess the potential carcinogenicity potential of the test material, SM-5688, [when administered] to rats by continuous dietary administration." (page 13 of 1991 rat report) The mouse report uses the same expression for mice. (page 12 of 1991 mouse report). Both studies were reanalyzed for Chelsea in 2011. Results from these reanalyses were were summarized in futher reports, both completed in 2011.

1.1. Conclusions and Recommendations

.

For both species, treatment was administered by dietary admixture, so assessing actual dose may be difficult. Note, as discussed in Section 1.3.1.1 below, rats were multiply housed together. This may cause problems in interpreting results.

Table 1. Design of Albino Rat Study

Treatment Group	# Animals	Dosage (mg/kg/day)
1. Control[1]	50	0
2. Low	50	10
3. Medium	50	30
4. High	50	100

[1]No treatment control group

Table 2. Design of Mouse Study

Treatment Group	# Animals	Dosage (mg/kg/day)
1. Control[1]	52	0
2. Low	52	30
3. Low-medium	52	100
4. Medium	52	300
5. High	52	1000

[1]No treatment control group

More detailed descriptions of the studies are provided in Section 3.2.1 and 3.2.2 below. In this report the treated actual treated groups, as opposed to the untreated controls, are sometimes referred to as "actual dose groups." Simple summary mortality tables are presented in the FDA analysis associated with these sections of the report.

In Appendix 1, Figures A.1.1 and A.1.2 for rats and Figures A.1.3 and A.1.4 for mice display Kaplan-Meier estimated survival curves for each study group for each species and gender

combination. The results of the tests of trend and differences in survival are displayed in Tables 3 and 4 below:

Table 3. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Rat Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rat Homogeneity over Groups 1-4	0.0584	0.0456	0.7221	0.9292
No trend over Groups 1-4	0.0598	0.0442	0.2977	0.5369
No Difference Between Groups 1 vs 4	0.1216	0.1128	0.3628	0.6936

From Figure A.1.1, in male rats, there appears to be a general tendency for the low dose group to have the highest survival, followed by the no treatment control, group 1, with the medium and high dose group fairly closely intertwined with the lowest survival. This is sufficient to result in significance levels close to the usual 0.05 level for the tests of homogeneity (logrank p = 0.0584, Wilcoxon p = 0.0456), and the tests of trend (logrank p = 0.0598, Wilcoxon p = 0.0442). However, although the survival curve of the control is generally above that of the high dose group, it is not sufficient to result in a statistically significant difference (logrank p=0.1216, Wilcoxon p= 0.1128). From Figure A.1.2, in female rats, for a brief period during the middle of the study, the medium dose group has the lowest survival, but none of the tests of differences in survival are statistically significamt at anything close to the usual level of significance (i.e. all six $p \geq 0.2977$)

Table 4. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Mouse Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxo	Log rank	Wilcoxon
Mice Homogeneity over Groups 1-5	0.0153	0.0117	0.0003	< 0.0001
No trend over Groups 1-5	0.0014	0.0011	0.0014	0.0005
No Difference Between Groups 1 vs 5	0.0035	0.0030	0.0010	0.0006

As with rats, figures A.1.3 and A.1.4, in Appendix 1, provide Kaplan-Meier survival curves for each mouse gender. In male mice, statrting roughly mid study or so, there is a clear simple dose related increase in mortality over increasing doses, consistent with the statistical tests above (all six $p \leq 0.0153$). Things are bit more complicated in female mice, in that the control generally has the lowest mortality, with the low and low-medium dose groups next in mortality, but close to each other, and the medium and high doses having the highest mortality. Again this is consistent with the results of the tests above (all six $p \leq 0.0013$).

Of course in a carcinogenicity study, primary interest is on the occurrence of cancers (see Sections 1.3.1.4 and 1.3.1.5). The statistical analysis is based on a comparison of tumor incidence over dose groups. Tables 5 and 6, below, display those tumor by organ combinations that had at least one test of trend or test of pairwise differences between an actual dose group and the untreated control that achieved a statistical significance level at the usual 0.05 level. For

each species by gender by organ the number of animals analyzed and used in the statistical tests is presented first. The tumor incidence for each organ is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high, medium, (plus low-medium in mice), and low dose groups.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.5, below, are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors (incidence > 1%) and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. As also discussed in section 1.3.1.5, using these adjustments for other tests can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Table 5. Potentially Statistically Significant Neoplasms in Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
Male Rats								
PANCREAS								
ISLET CELL ADENOMA	7	17	14	12	.4375	.1540	.0698	.0169
Islet Adenoma/Carcinoma	9	19	18	14	.4572	.1710	.0352	.0220
Female Rats								
SUBCUTIS								
FIBROMA	0	1	4	1	.5547	.5000	.0256	.3333
Fibroma/Fibrosarcoma	0	1	5	1	.5727	.5000	.0070	.3333

Using the Haseman-Lin adjustment for multiplicity, since they would be classified as common tumors, in male rats the test of differences between the low dose and control groups in islet cell adenoma was not significant (i.e. $p = 0.0169 > 0.01$), nor were the tests between the low and medium dose groups in pooled islet cell adenoma/carcinoma (both $p = 0.0352, 0.0220 > 0.01$, respectively). From the incidence in the control group in female rats, we would define fibromas and pooled fibroma/fibrosarcomas as rare tumors. If one then accepts the increase in overall type I error to some value above the rough 10% level inherent in testing the medium and low dose groups, the pairwise tests between the medium dose group and controls in fibroma was statistically significant ($p = 0.0256 \leq 0.05$). Similarly, the pairwise test between the medium dose group and controls in pooled fibroma and fobrosarcoma would be statistically significant ($p = 0.007 \leq 0.05$). No other tests achieved the usual 0.05 level of significance.

In mice, those organ-tumor combinations with at least one nominally statistically significant result at the typical level ($p \leq 0.05$) in mice are summarized below. The period '.' in these tables denotes the p-values of tests of dose groups with no tumors in any group. Those tables are included in Appendix 3.

Table 6. Potentially Statistically Significant Neoplasms in Mice

Organ / Tumor	Incidence Cntrl	Low	LwMd	Med	Hi	Significcance Levels Trend	HvsC	MvsC	LMvsC	LvsC
Male Mice										
TESTES										
# Evaluated	52	15	11	18	52					
INTERSTITIAL CELL TUMOUR	0	2	0	1	0	.7847	.	.	.2571	.0475
Female Mice										
LUNGS + BRONCHI										
# Evaluated	52	26	19	27	52					
PULMONARY ADENOMA	3	6	3	3	4	.7810	.5000	.1898	.3324	.0331
UTERUS										
# Evaluated	52	48	48	45	52					
Endo. Adenocarc./Strom. Sarcoma	0	0	0	1	2	.0338	.2476	.	.4639	.

Again, accepting the inflated type I error, the test of differences between the low dose and controls in interstitial cell tumor of the testes in male mice was barely statistically significant ($p = 0.0475 < 0.05$), since it is classified as a rare tumor. However, the test in pulmonary adenoma in female mice was not significant ($p = 0.0331 > 0.01$) since it is not classified as a rare tumor. Even though classified as a rare tumor, the test of trend in pooled endometrial adenocarcinoma and stromal sarcoma was not statistically significant ($p = 0.0338 > 0.025$). No other tests in mice even achieved the nominal 0.05 level.

Complete incidence tables in both species are provided in tables A.2.3 through A.2.6 in Appendix 2.

1.2. Brief Overview of the Studies

This submission had a rat study:

Study SUP1/90147 SM-5688 Potential Tumorigenic Effects in Prolonged Dietary Administrationto Rats,

and a very similar, mouse study:

Study SUP4/90180 SM-5688 Potential Tumorigenic Effects in Dietary Administration to Mice for 80 Weeks,

Both studies were conducted in (b) (4) A reanalysis of the orginal data was conducted by (b) (4) in 2011. Somewhat detailed descriptions of these studies are available in Sections 3.2.1 and 3.2.2, below.

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

1.3.1.1. Housing of Animals:

In the rat study animals were housed five per cage while in mice were housed individually. Since dosing was dietary admixture, with multiple animals per cage it is difficult to equalize dosage across animals within a cage, and thus insure that animals within a cage have the same dose. A further problem is that it constitutes a restriction on the randomization, thus invalidating any randomization justification of the analysis. In general, multiple housing of animals may cause other statistical problems in the analysis. It is possible that proximity within cages, including attendant fighting, may induce positive correlations in response, while within cage competion for food possibly could induce negative correlations. Variations in dose within cage groups would tend to produce positive correlations. Thus, because of this housing, the within treatment estimated variances may be too large or too small, resulting in conservative or anti-conservative tests (in terms of Type I error). Unless it has been clearly shown that tumor incidence is independent of cage, from a purely statistical analysis point of view, this reviewer would generally recommend single housing of animals. Without data on the actual caging these effects in the rat study can not be investigated. Again, such concerns do not apply to the mouse study.

1.3.1.2. Survival Analysis:

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

1.3.1.3. Multiplicity of Tests on Survival:

Using the logrank and Wilcoxon tests, for each gender in rats and mice there are at least six tests of survival differences in each gender. If we were to assume that any set of tests are independent across comparisons, which clearly they are not, and assume that there is absolutely no difference in survival, the probability of at least one statistically significant result in each gender in each species, at the usual 0.05 level, is about 0.2649. Under these conditions, the probability of at least one significant comparison within each species was .4596. This gives at

least some measure of the possible price paid for the multiplicity of hypothesis tests in the frequentist paradigm.

1.3.1.4. Tests on Neoplasms:

The Sponsor's analyses use Peto analyses of neoplasms. The analyses in this report are based on poly-k analysis of tumor incidence. The poly-k test is a modification of the original Cochran-Armitage test of trend in response to dose, adjusted for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). Roughly, animals that die early without the tumor under consideration are down-weighted by an amount proportional to the kth power of the ratio of the animal's lifetime to the length of the study. It was noted in the report of the Society of Toxicological Pathology "town hall" meeting in June 2001 that this poly-k modification of the Cochran-Armitage tests of trend has been recommended over the corresponding Peto tests.

1.3.1.5. Multiplicity of Tests on Neoplasms:

Frequentist hypothesis testing involves accepting or rejecting hypotheses about the parameters of interest on the basis of the values of some statistic. If one does not provide some sort of multiplicity adjustment to the significance level, the chances of rejecting one or more true null hypothesis increases as the number of such tests increases. To avoid this, it is common to adjust for multiplicity in hypothesis testing resulting in an adjustment in experiment-wise Type I error (i.e., the probability of rejecting a true null hypothesis). Based on his extensive experience with such carcinogenicity analyses in standard laboratory rodents, 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. Similarly, Lin and Rahman (1998) showed that tests of trend should be tested at a 0.025 level for rare tumors and 0.005 for common tumors. This approach is intended to balance both Type I error and Type II error (i.e., the error of concluding there is no evidence of a relation to tumorgenicity when there actually is such a relation).

Significance levels of the pairwise tests between the untreated control and the low, low-medium (in mice), and medium dose groups are also provided. Including these tests can be expected to increase the overall type I error rate to some level above the rough 10% level specified above. Even if one uses the Haseman-Lin-Rahman rules, the overall type I error associated with including the tests between the control and the low, low-medium, and medium dose groups may be considerably larger than the rough 10% when these rules are restricted to the test of trend and pairwise differences between the high dose and the no treatment control.

1.3.1.6. Validity of the Designs:

When determining the validity of designs there are two key points:

1) adequate drug exposure,
2) tumor challenge to the tested animals.

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

designed to achieve the greatest likelihood of tumorigenicity. This reviewer would suggest that the 80-week mouse study may have been terminated too early to establish adequate exposure.

Lin and Ali (2006), quoting work by Haseman, have suggested that in standard laboratory rodent species, a survival rate of about 25 animals, out of 50 or more animals, 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 that as a percentage of animals that survived to week 91, this criterion is clearly met in rats, but is exceded in mice. (Please see Tables 13 and 14 on pages 14 and 15, and Tables 19 and 20 on page 19). Like the other comments in this section this requires the expertise of the toxicologist, but may suggest that the MTD was was met in rats, while the length of the mouse study may have been too short to draw a similar conclusion.

The mean weight values and derived differences and ratios in the following table were taken directly from the Sponsor's reports (Rat Tables 2a and 2b, pages 23-27, Mouse Table 3, pages 34-37). In the mouse table below, Table 8, the change from baseline in the table below is the simple difference between the means at the specified dates, and thus animals that die are only in included in the mean at study initiation, not at the end of the study.

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

Dose Group	Dose mg/kg/ day	Males				Females			
		Week		Change from Baseline	% change relative to control	Week		Change from Baseline	% change relative to control
		0	104			1	104		
1. Control	0	173	795	621		142	533	392	
2. Low	10	173	788	615	98%	143	552	411	105%
3. Medium	30	174	777	604	97%	142	537	395	101%
4. High	100	174	790	616	99%	142	517	375	96%

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

Dose Group	Dose mg/kg/ day	Males				Females			
		Week		Change from Baseline	% change relative to control	Week		Cha nge from Baseline	% change relative to control
		0	80			0	80		
1. Control	0	25	47	22		21	39	18	
2. Low	30	24	47	23	106%	21	38	17	94%
3. Low- Med.	100	25	46	21	95%	21	39	18	100%
4. Medium	300	25	44	19	86%	20	38	18	100%
5. High	1000	25	43	18	82%	21	38	17	94%

Chu, Ceuto, and Ward (1981), citing earlier work by Sontag *et al* (1976) recommend that the MTD "is taken as 'the highest dose that causes no more than a 10% weight decrement as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span' " From Table 7 above, no rat dose

groups exceeded the weight decrement criterion. However, even with the foreshortend mouse studies, the weight criterion was exceeded in the medium and high dose groups in male mice. Although this requires the expertise of the toxicologist, this may be evidence that the MTD was exceeded in male mice.

In rats, the Sponsor summarizes the dose effect on body weight and food consumption as: "There was no apparent effect of treatment on these parameters" (page 11 of 1991 rat report) In mice the Sponsor summarizes results as "Slightly lower mean bodyweight gains were evident from Week 10 to Week 60 for males receiving 300 or 1000 mg/kg/day when compared with controls Group mean bodyweight gains for all other treated mice were considered comparable throughout the study" (page 8 of 1991 mouse report) Further, "Slightly lower cumulative mean food intake was evident for males receiving 300 or 1000 mg/kg/day when compared to control values from Week 10 to Week 60 of the study. Group mean cumulative food intake for all other treated mice was considered comparable with controls throughout the study" (page 9 of 1991 mouse report)

Again from 2) above, excess mortality not associated with any tumor or sacrifice in the higher dose groups might suggest that the MTD was exceeded. This suggests that a useful way to assess whether or not the MTD was achieved is to measure early mortality not associated with any identified tumor. If this is high in the higher dose groups it suggests that animals tend to die before having time to develop tumors. Tables 9 and 10, below, display the number of animals in each dose group that died of a natural death or moribund sacrifice, but did not show any tumors (i.e., the "Event"):

Table 9. Natural Death with No Identified Tumor in Rats (Male/Female)

	1. Control	2. Low	3. Medium	4. High
Males Event	3	4	6	6
No event	47	46	44	44
Females Event	0	1	1	3
No event	50	49	49	47

In rats there was no particular evidence of heterogeneity over dose group in early death without tumors (Males: Logrank p = 0.5366, Wilcoxon p = 0.4369, Females: Logrank p = 0.2671, Wilcoxon p = 0.2458).

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

	1. Control	2. Low	3. Low	4.Medium	5. High
Males Event	3	7	8	15	14
No event	49	45	44	37	38
Females Event	8	9	11	19	17
No event	44	43	41	33	35

However, in both genders in mice there is clear evidence of a difference in early death without tumors. Adjusting for time to event, all tests were statistically significant (Males

Overall: Logrank p = 0.0087, Wilcoxon p = 0.0079, Control versus High: Logrank p = 0.0032, Wilcoxon p = 0.0035, Females Overall: Logrank p = 0.0036, Wilcoxon p = 0.0035, Control versus High: Logrank p = 0.0082, Wilcoxon p = 0.0050). Like the other observations above, these require the expertise of the toxicologist, but these tests seem to provide evidence that not only was the mouse study terminated too early, but the MTD was exceeded in both mouse genders.

1.3.2. Statistical Findings

Please see Section 1.1 above.

2. INTRODUCTION

2.1. Overview

This submission summarizes the results of a two year rat and an 18 month mouse study to assess the carcinogenic potential of droxidopa when dosed with daily dietary administartion. Both studies were conducted in the late 1980's by (b) (4) at their research center in (b) (4)

2.2. Data Sources

The Sponsor provided two SAS transport files, labeled ea-b-2-1-tumor.xpt for rats and ea-b-1-1-tumor.xpt for mice, where export file was reformatted to a SAS tumor data sets named tumor.sas7bdat. These largely following the standard specifications of the FDA Biometrics requested format.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1. Study SUP1/90147 SM-5688 Potential Tumorigenic Effects in Prolonged Dietary Administrationto Rats,

STUDY DURATION: 104 Weeks
EXPERIMENTAL (DOSING) START DATE: 23 December 1987
DOSING TERMINATION: 20 December 1989
RAT STRAIN: Sprague Dawley CD® Rats
ROUTE: Dietary admixture

The basic design of the rat study has three drug dosing groups, and an untreated control, as summarized in Table 11, below, actually a repeat of Table 1:

Table 11. Design of Rat Study

Treatment Group[1]	# Animals	Dosage (mg/kg/day)
1. Control	50	0
2. Low	50	10
3. Medium	50	30
4. High	50	100

[1] No treatment control group

The Sponsor summarized justified the study design as follows: "The dosage levels used in this study were chosen by the Sponsor based on previous work on the test masterial (HRC Report No. SMO 276/871584) and toxicity studies conducted by the Sponsor. The test species was chosen according to regulatory requirements and the Sprague-Dawley (CD) strain was chosen due to the availability of background data at this laboratory. The dietary route was chosen as it is the anticipated route of human exposure to the test material" (page 1 of 1991 rat report)

"The rats were stratified by the bodyweight and allocated to each group with approximate equal average bodyweight before commencement of treatment." (page 3 of 1991 report)

"All rats were allowed free access to tap water and SDS Rat and Mouse No. 1 modified maintenance diet (powdered). There was no information available to the Study Director to indicate that any non-nutrient substance likely to influence the effect of the test material could reasonably be expected to be present in the diet . . ." (page 2 of 1991 rat report)

"Before the strart of treatment the proposed diet mixing procedures were checked by chemical analysis of trial diets to confirm that the proposed procedures produced homogeneous diet, that the accuracy of mixing was acceptable and that the concentration of test material in the diet remained unchanged between preparation and administration" (page 4 of 1991 report)

Animals were housed together in groups of five animals. As discussed in Section 1.3.1.1, this may cause problems in specifying an appropriate analysis and is not generally recommended, particularly with dietary dosing.

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 rats.

Sponsor's Survival analysis:

The Sponsor provided the following table showing the percentage of animals surviving at termination. The Sponsor noted that: "The mortality distribution during the 104-week treatment period was as follows:

[**Table 12. Sponsor Summary on Mortality**]

Group and sex	1♂	2♂	3♂	4♂	1♀	2♀	3♀	4♀
Dosage (mg/kg/day)	Control	10	30	100	Control	10	30	100
Incidence	24	23	32	31	28	28	24	22
% mortality	48	46	64	62	56	56	48	44
% survival	52	54	36	38	44	44	52	56

In addition one animal from each of Groups 1♂, 2♂, .3♂, 1♀, and 4♀ died during the termination period." (page 10 of 1991 rat report)

"The overall (Weeks 1 to 104) mortality incidence of males receiving 30 or 100 mg/kg/day was slightly higher than that of control. The statistical analysis test for trend was significant (P = 0.033) for the male data, but pairwise comparisons did not attain formal significance. There were no ante-mortem clinical findings indicative of treatement." (page 11 of 1991 rat report)

Sponsor's Tumorigenicity analysis:

The Sponsor used a standard Peto style analysis of carcinogenicity, with the Haseman-Lin adjustment for multiplicity (please see Section's 1.3.1.5 and 1.3.1.6). These results of this analysis were summarized results as follows: "The results of the statistical analysis of tumour incidence confirm the interpretation of the original study report that no treatment relationship was seen in the incidence or distribution of neoplastic findings in this study. Statistically significant differences were seen in the occurrence of tumours of the pancreas and thyroid in male animals, but these were not considered to be related to treatment with SM-5688.

"Pancreas
Statistically significant increases were seen in the incidence of benign Islet cell adenoma or benign and malignant Islet cell tumours combined in males of the low and intermediate dosage groups in pairwise comparisons with the Control group. However, the incidence of these findings in males of the high dose group was not statistically significant; the trend test was not significant; and there was no clear dose-relationship in the incidence of these findings. In addition, there was no effect of treatment on the incidence of Islet cell hyperplasia in male animals, which would be expected if a treatment-related effect on proliferative lesions of the Islet cells was suspected.

"Thyroids
A statistically significant trend test was found for parafollicular cell carcinoma in males. None of the pair-wise comparisons were statistically significant however; there was no clear dose-relationship in the incidence of this finding, and no effect of treatment on the incidence of parafollicular cell hyperplasia or benign parafollicular cell tumours in males.
Because proliferative findings observed in endocrine tissues represent a continuum of change from hyperplasia to neoplasia, the lack of an effect of treatment on the incidence of the relevant hyperplastic changes adds weight to the interpretation that the neoplastic findings seen are not related to treatment, and the statistically significant differences observed have arisen by chance." (pages 11-12 of 2011 rat report)

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following tables (Table 13 for male rats, Table 14 for females) summarize the mortality results for the study groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses.

Table 13. Summary of Male Rats Survival (dosed at estimated mg/kg/day)

Period (Weeks)	Control veh ~ 0	Low 10 mg	Medium 30 mg	High 100 mg
1-52	0/50[1] 100%[2]	0/50 100%	2/50 96%	1/50 98%
53-78	5/50 90%	3/50 94%	7/48 82%	9/49 80%
79-93	10/45 70%	7/47 80%	11/41 60%	11/40 58%
94-104	10/35 42%	14/40 28%	13/30 34%	10/29 38%
Terminal [3] 105	25	26	17	19

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

Table 14. Summary of Female Rat Survival (dosed at estimated mg/kg/day)

Period (Weeks)	Control veh ~ 0	Low 10 mg	Medium 30 mg	High 100 mg
1-52	0/50[1] 100%[2]	1/50 98%	1/50 98%	2/50 96%
53-78	5/50 90%	5/49 88%	8/49 82%	4/48 88%
79-93	8/45 74%	7/44 74%	6/41 88%	10/44 68%
94-104	16/37 42%	15/37 44%	9/35 52%	7/34 54%
Terminal [3] 105	21	22	26	27

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

Table 15 below provides the significance levels of the tests of homogeneity and trend over dose groups as proposed in Section 1.3.1.1, above.

Table 15. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Rat Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rat Homogeneity over Groups 1-4	0.0584	0.0456	0.7221	0.9292
No trend over Groups 1-4	0.0598	0.0442	0.2977	0.5369
No Difference Between Groups 1 vs 4	0.1216	0.1128	0.3628	0.6936

From Figure A.1.1, in Appendix 1, in male rats there appears to be a general tendency for the low dose group to have the highest survival, followed by the control group 1, with the medium and high dose group to be closely intertwined and with the lowest survival. This is sufficient to result in close to the usual 0.05 level of statistical significant for the tests of homogeneity (logrank p = 0.0584, Wilcoxon p = 0.0456), and the tests of trend (logrank p = 0.0598, Wilcoxon p = 0.0442). However, although the survival curve of the control is generally above that of the high dose group, it is not sufficient to result in a statistically significant difference (logrank p=0.1216, Wilcoxon p= 0.1128). From Figure A.1.2, in female rats, for a brief period during the middle of the study the medium dose group has the lowest survival, but none of the tests of differences in survival are statistically significamt at anything close to the usual level of significance (i.e. all six $p \geq 0.2977$)

Tumorigenicity analysis:

As discussed in Section 1.3.1.5, the Haseman-Lin-Rahman rules for adjusting for multiplicity, specify that for a rough 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. Similarly, Lin and Rahman (1998) showed that tests of trend should be tested at a 0.025 level for rare tumors and 0.005 for common tumors.

Table 16. Potentially Statistically Significant Neoplasms in Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
Male Rats								
PANCREAS								
ISLET CELL ADENOMA	7	17	14	12	.4375	.1540	.0698	.0169
Islet Adenoma/Carcinoma	9	19	18	14	.4572	.1710	.0352	.0220
Female Rats								
SUBCUTIS								
FIBROMA	0	1	4	1	.5547	.5000	.0256	.3333
Fibroma/Fibrosarcoma	0	1	5	1	.5727	.5000	.0070	.3333

Using the Haseman-Lin adjustment for multiplicity, in male rats the test of differences between the low dose and control groups in islet cell adenoma was not significant (i.e. $p = 0.0169 > 0.01$), nor were the tests between the low and medium dose groups in pooled islet cell adenoma/carcinoma (both $p = 0.0352, 0.0220 > 0.01$, respectively). From the incidence in the control group in female rats, we would define fibromas and pooled fibroma/fibrosarcomas as rare tumors. If one then accepts the increase in overall type I error to some value above the rough 10% level inherent in testing the medium and low dose groups, the pairwise test between the medium dose group and controls in fibroma was statistically significant ($p = 0.0256 \leq 0.05$). Similarly, the pairwise test between the medium dose group and controls in pooled fibroma and fibrosarcoma would be statistically significant ($p = 0.007 \leq 0.05$). No other tests achieved the 0.05 level of significance.

Complete incidence tables are provided in tables A.2.3 and A.2.4 of Appendix 2.

3.2.2. Study SUP4/90180 SM-5688 Potential Tumorigenic Effects in Dietary Administration to Mice for 80 Weeks,

STUDY DURATION: 80 weeks
EXPERIMENTAL START DATE (INITIATING DOSING): 16 May 1988
DOSING TERMINATION: 17 December 1989
MOUSE STRAIN: Charles River CD-1 mice
ROUTE: Daily dietary admixture

Gross aspects of the study designs for the main study animals are summarized below (a repeat of table 2) :

Table 17. Design of Mouse Study

Treatment Group	# Animals	Dosage (mg/kg/day)
1. Control[1]	52	0
2. Low	52	30
3. Low-medium	52	100
4. Medium	52	300
5. High	52	1000

[1] No treatment control group

The Sponsor summarized study conduct as follows: "The objective of this study, performed at the (b) (4) was to assess the potential tumorigenicity of the test material, SM-5688, to mice by continuous dietary administration for at least 80 weeks.

"This study was designed in accordance with toxicity test guidelines published in Notification No. 118 of the Pharmaceutical Affairs Bureau, Ministry of Health and Welfare dated February 15, 1984.

"The mouse was the species of choice due to regulatory requirements and the strain chosen on account of the availability of background data at this laboratory. The dietary route of administration was chosen because the anticipated route of clinical administration is oral.

"The treatment levels of 30, 100, 300 and 1000 mg/kg/day used in this study were chosen by the Sponsor with reference to a 13-week preliminary study (SMO 255/871087) performed at (b) (4) In this preliminary study dietary inclusion levels equivalent to 3000 mg/kg/day induced toxicity manifest as myocardial fibrosis and basophilic cortical tubules of the kidneys, and other minor changes associated with renal disfunction, in a proportion of males and females. One female treated at 1000 mg/kg/day also showed focal myocardial fibrosis." (page 1 of 1991 mouse 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.

Sponsor's Survival analyisis:

The Sponsor summarizes mortality by noting that: "There was a total of 150 deaths during the study. The distribution of mortalities was as follows:

[**Table 18. Sponsor Summary on Mortality**]

Weeks	Group and dosage (mg/kg/day)									
	1♂ Control	2♂ 10	3♂ 30	4♂ 100	5♂ 1000	1♀ Control	2♀ 10	3♀ 30	4♀ 100	5♀ 1000
1 – 80	8	11	12	17	21	5	11	14	24	20
81 - 83	0	0	0	1	0	3	2	0	0	1
% survival at Termination	85	79	77	65	60	85	75	73	54	60

"Statistical analysis revealed mortality at 300 and 1000 mg/kg/day to be significantly increased for both sexes in comparison with concurrent controls. These increased incidences were principally attributable to an exacerbation of amyloidosis, a spontaneous age-related change in this strain of laboratory mouse. This was not seen to a similar degree at 30 or 100 mg/kg/day. Mortality amongst females treated with 100 mg/kg/day was also significantly increased in comparison with control." (page 11 of 1991 rat report)

Sponsor's Tumorigenicity analysis:

The Sponsor used a standard Peto style analysis of carcinogenicity, with the Haseman-Lin adjustment for multiplicity. The results of this analysis were summarized as follows: "The results of the statistical analysis of tumour incidence confirm the interpretation of the original study report that no treatment relationship was seen in the incidence or distribution of neoplastic findings in this study." (page 8 of 2011 mice report)

3.2.1.2. FDA Reviewer's Results

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

Survival analysis:

The following tables (Table 19 for male mice, Table 20 for females) summarize the mortality results for the study groups. The data were grouped for the specified time period, and present the number of deaths during the time interval over the number at risk at the beginning of the interval. The percentage cited is the percent that survived at the end of the interval. In these tables the terminal period only includes those animals were sacrificed. Animals that died of other causes during the terminal period are included in the preceding, but overlapping time period. The Kaplan-Meier survival plots in Appendix 1 provide a more detailed picture of the profile of mortality losses.

Table 19. Summary of Male Mice Survival (dosed at estimated mg/kg/day)

(Weeks)	Veh Sim. ~ 0	Low.	Low-Med 25 mg	Medium 75 mg	High 200 mg
1-40	1/52[1] 98,.1%[2]	2/52 96.2%	0/52 100%	1/52 98,.1%	2/52 96.2%
41-60	2/51 94.2%	0/50 96.2%	2/52 96,.2%	4/51 90.4%	7/50 82. 7%
61-83	5/49 84.6%	9/50 78.8%	9/50 76.9%	13/47 65.4%	12/43 59.6%
Terminal[3] 83-84	44	41	41	34	31

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

Table 20. Summary of Female Mice Survival (dosed at estimated mg/kg/day)

(Weeks)	Veh Sim. ~ 0	Low.	Low-Med 25 mg	Medium 75 mg	High 200 mg
1-40	0/52[1] 100%[2]	0/52 100%	2/52 96.2%	2/52 96.2%	2/52 96.2%
41-60	1/52 98,.1%	1/52 98,.1%	1/52 94,.2%	11/50 75.0%	12/50 73.1%
61-83	7/51 84.6%	12/51 75.0%	11/49 73.1%	11/39 43.1%	7/38 43.1%
Terminal[3] 83-84	44	39	38	28	31

[1] number of deaths / number at risk
[2] overall per cent survival to end of period.
[3] number of animals that survived to terminal sacrifice

The following table, Table 21, summarizes the results from tests comparing survival profiles across study groups in the tumorigenicity data sets:

Table 21. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Mouse Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxo	Log rank	Wilcoxon
Mice Homogeneity over Groups 1-5	0.0153	0.0117	0.0003	< 0.0001
No trend over Groups 1-5	0.0014	0.0011	0.0014	0.0005
No Difference Between Groups 1 vs 5	0.0035	0.0030	0.0010	0.0006

As with rats, figures A.1.3 and A.1.4, in Appendix 1, provide Kaplan-Meier survival curves for each mouse gender. In male mice, statrting roughly mid study or so, there is a clear simple dose related increase in mortality over increasing doses, consistent with the statistical tests above (all six $p \leq 0.0153$). Things are bit more complicated in female mice, in that the

control generally has the lowest mortality, with the low and low-medium dose groups next in mortality, but close to each other, and the medium and high doses having the highest mortality. Again this is consistent with the results of the tests above (all six $p \leq 0.0013$).

Tumorigenicity analysis:

To reiterate, for a rough 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, while tests of trend should be tested at a 0.025 level for rare tumors and 0.005 for common tumors. Those organ-tumor combinations with at least nominally statistically significant result ($p \leq 0.05$) in mice are summarized below:

Table 22. Potentially Statistically Significant Neoplasms in Mice

Organ	Incidence					Significance Levels				
Tumor	Cntrl	Low	LwMd	Med	Hi	Trend	HvsC	MvsC	LMvsC	LvsC
Male Mice										
TESTES										
# Evaluated	52	15	11	18	52					
INTERSTITIAL CELL TUMOUR	0	2	0	1	0	.7847	.	.	.2571	.0475
Female Mice										
LUNGS + BRONCHI										
# Evaluated	52	26	19	27	52					
PULMONARY ADENOMA	3	6	3	3	4	.7810	.5000	.1898	.3324	.0331
UTERUS										
# Evaluated	52	48	48	45	52					
Endo. Adenocarc./Strom. Sarcoma	0	0	0	1	2	.0338	.2476	.	.4639	.

Again, with the inflated type I error, the tests of differences between the low dose and controls in interatitial cell tumor of the testes in male mice was barely statistically significant ($p = 0.0475 < 0.05$). while the test in pulmonary adenoma in female mice and was not significant ($p = 0.0331 > 0.01$), since it is not classified as a rare tumor. Further, the test of trend in pooled endometrial adenocarcinoma and stromal sarcoma was not statistically significant ($p = 0.0338 > 0.025$). No other tests in mice even achieved the nominal 0.05 level.

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. FDA Survival Analysis

Simple summary life tables in mortality are presented in the report (Tables 13, 14, 19, and 20, above). Kaplan-Meier estimated survival curves across study groups for each gender are displayed below in Figures A.1.1 and A.1.2 for rats and Figures A.1.3 and A.1.4 for mice. These plots include 95% confidence intervals around each survival curve (colored area around each curve). These plots are also supported by tests of homogeneity in survival over the different treatment groups, tests of trend in survival over increasing dose, and the results of pairwise comparisons between the high dose group and the control. The statistical significance levels (i.e., p-values) are provided in Tables A.1.1. and A.1.2., below. One might note that the log rank tests places greater weight on later events, while the Wilcoxon test tends to weight them more equally, and thus places more weight on earlier events than does the log rank test.

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

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Rat Homogeneity over Groups 1-4	0.0584	0.0456	0.7221	0.9292
No trend over Groups 1-4	0.0598	0.0442	0.2977	0.5369
No Difference Between Groups 1 vs 4	0.1216	0.1128	0.3628	0.6936

From Figure A.1.1, in male rats, there appears to be a general tendency for the low dose group to have the highest survival, followed by the control, group 1, with the medium and high dose group to be closely intertwined and with the lowest survival. This is sufficient to result in close to the usual 0.05 level of statistical significant for the tests of homogeneity (logrank p = 0.0584, Wilcoxon p = 0.0456), and the tests of trend (logrank p = 0.0598, Wilcoxon p = 0.0442). However, although the survival curve of the control is generally above that of the high dose group, it is not sufficient to result in a statistically significant difference (logrank p=0.1216, Wilcoxon p= 0.1128). From Figure A.1.2, in female rats, for a brief period during the middle of the study the medium dose group has the lowest survival, but none of the tests of differences in survival are statistically significamt at anything close to the usual level of significance (i.e. all six $p \geq 0.2977$)

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

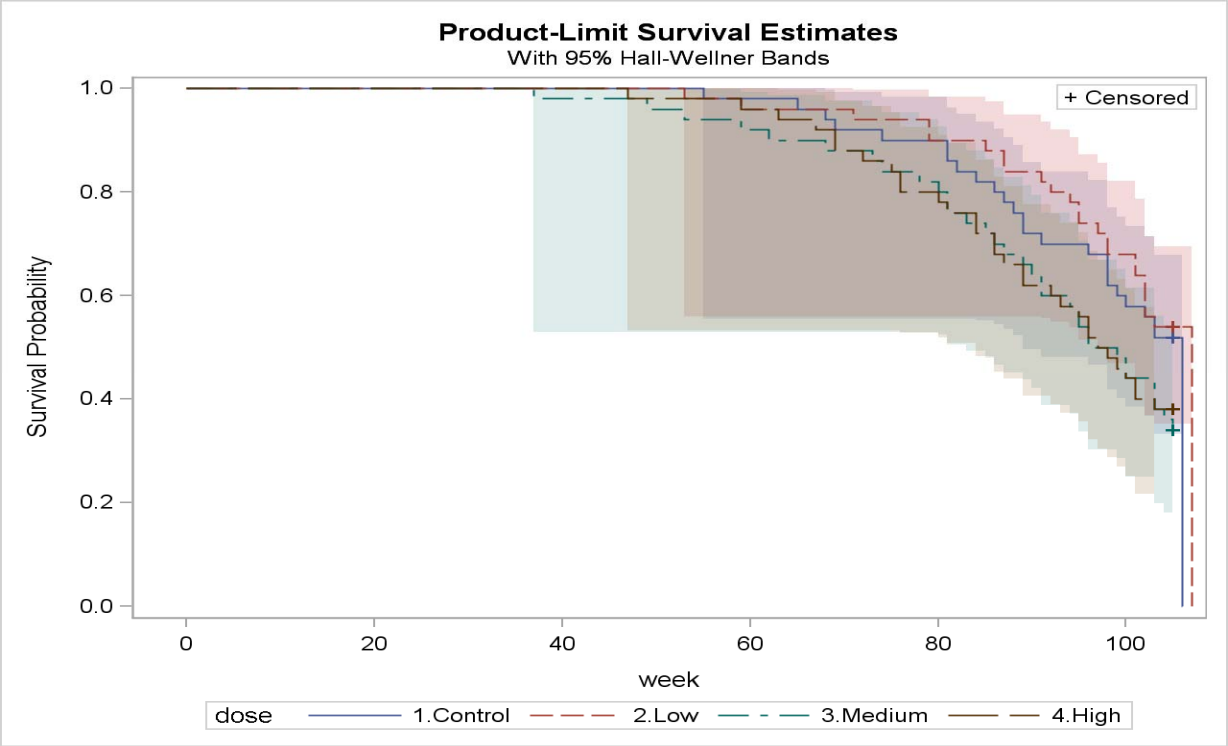
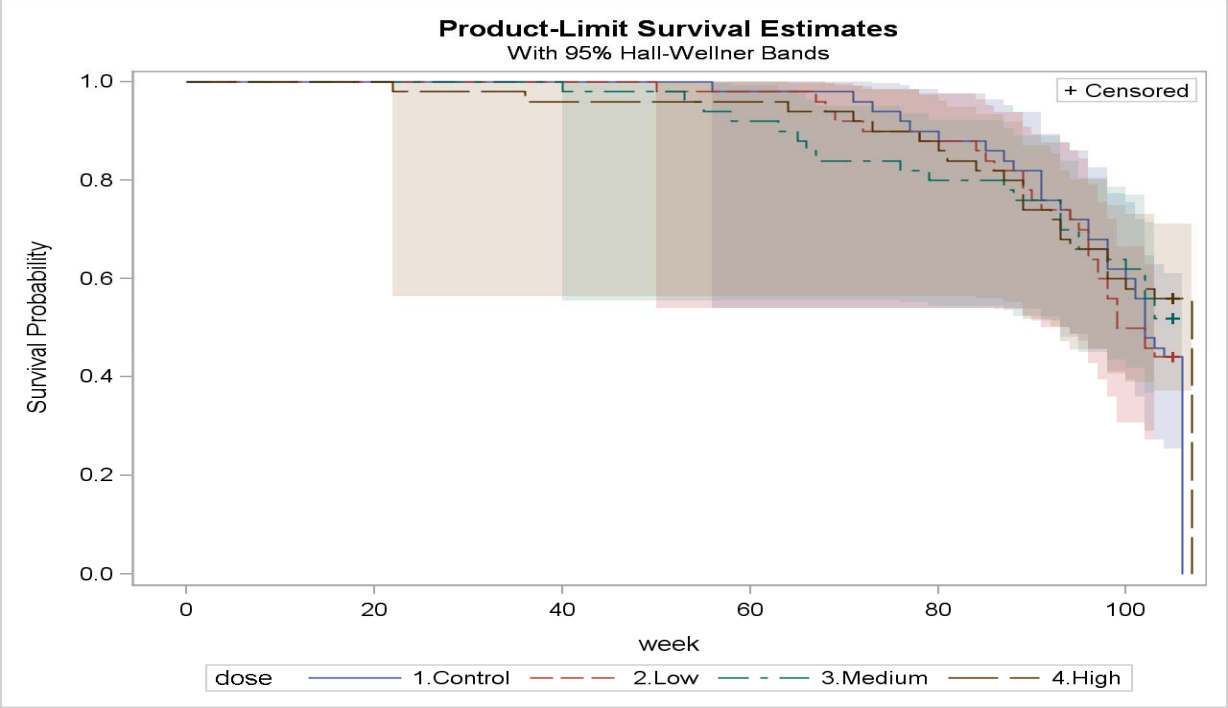


Figure A.1.2 Kaplan-Meier Survival Curves for Female Rats



Results for mice are presented below:

Table A.1.2. Statistical Significances of Tests of Homogeneity and Trend in Survival in the Mouse Study

Hypothesis Tested	Males		Females	
	Log rank	Wilcoxon	Log rank	Wilcoxon
Mice Homogeneity over Groups 1-4	0.0153	0.0117	0.0003	< 0.0001
No trend over Groups 1-4	0.0014	0.0011	0.0014	0.0005
No Difference Between Groups 1 vs 4	0.0035	0.0030	0.0010	0.0006

Figures A.1.3 through A.1.4, below, provide similar survival curves for each mouse gender. In male mice, statrting roughly mid study or so, there is a clear simple dose related increase in mortality over increasing doses, consistent with the statistical tests above (all six $p \leq 0.0153$). Things are bit more complicated in female mice, in that the control generally has the lowest mortality, with the low and low-medium dose groups next in mortality, but close to each other, and the medium and high doses having the highest mortality. Again this is consistent with the results of the tests above (all six $p \leq 0.0013$).

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

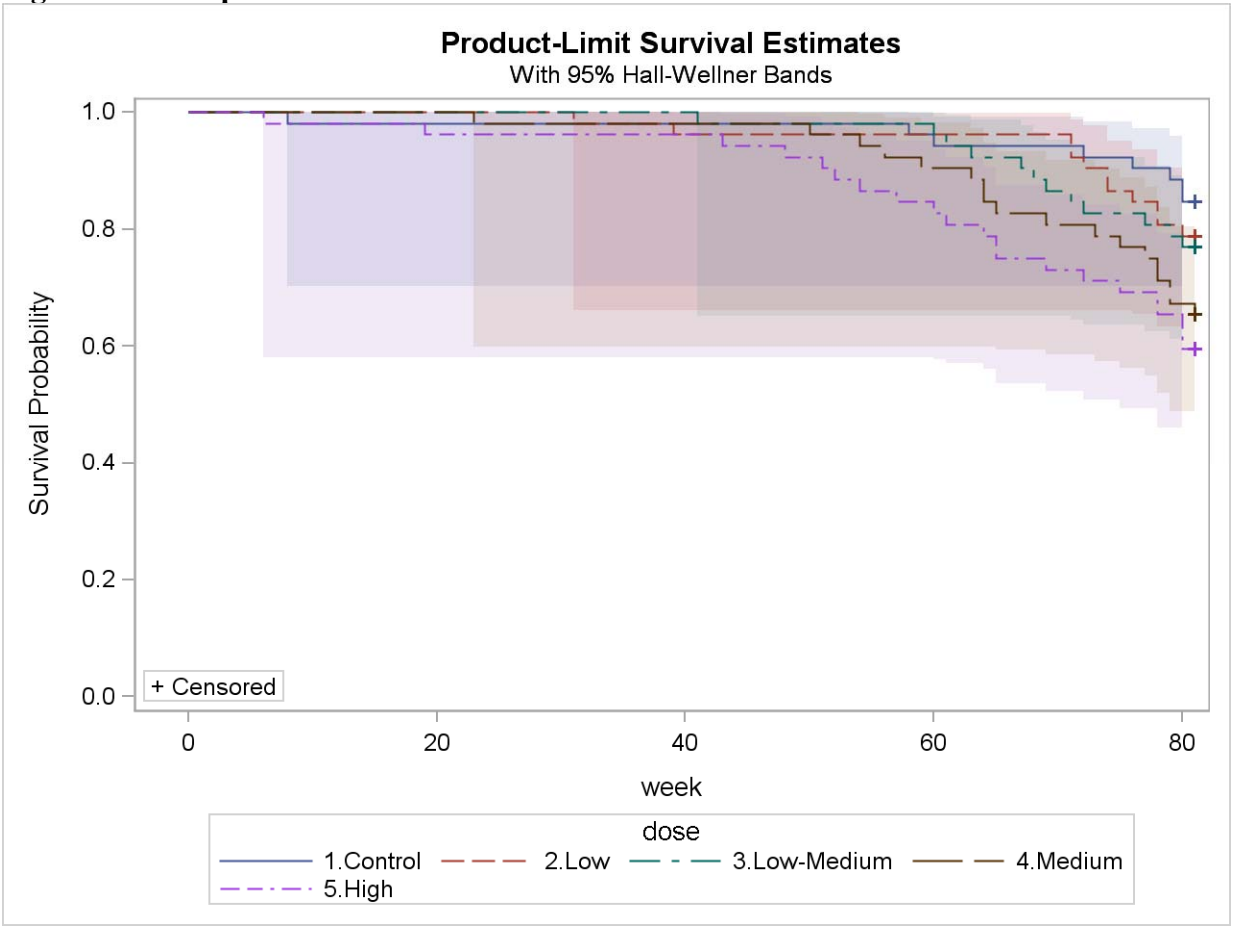
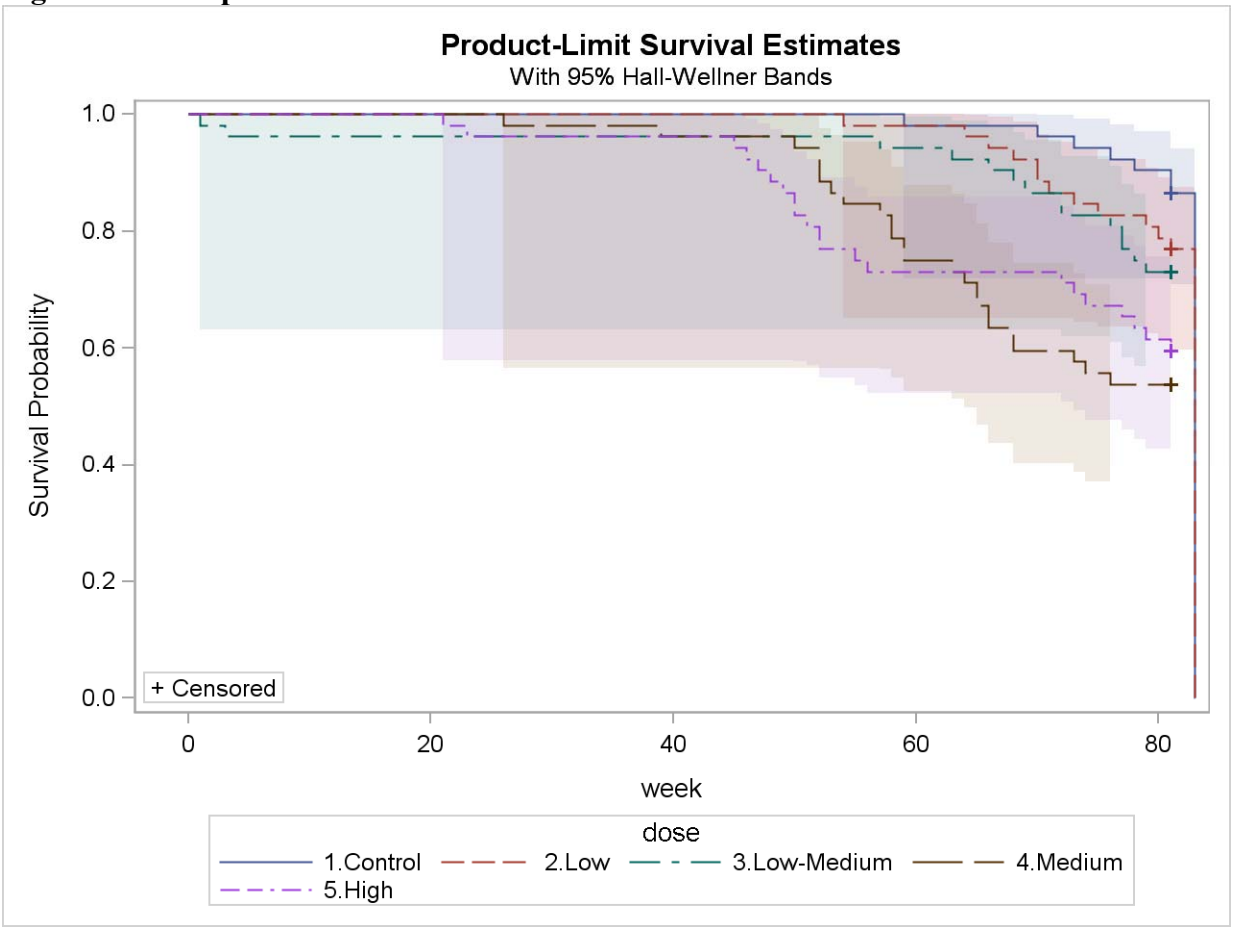


Figure A.1.4 Kaplan-Meier Survival Curves for Female Mice



Although the dose response is a lttle more complicated in female mice than in male mice, the survival curves are more separated, thus explaining the greater significance levels in the tests for female mice than in male mice.

Appendix 2. FDA Poly-k Tumorigenicity Analysis

The poly-k test, here with k=3, modifies the original Cochran-Armitage test to adjust for differences in mortality (please see Bailer & Portier, 1988, Bieler & Williams, 1993). The tests used here are small sample exact permutation tests of tumor incidence. These do assume all marginal totals are fixed, a debatable assumption. This assumption implies that in the pairwise tests when one dose group has no tumors of the specific type and the other does, there is only one permutation of this pattern. Since that means that the only permutation of the data is the one observed, that means that all possible permutations are as extreme as the pattern observed, and thus the significance level of the observed pattern can be logically expressed as 1.0. One could use the same sort of argument when there were no tumors of the specific type being analyzed in either column of the 2x2 table corresponding to a pairwise comparison. Then an argument could be made that the p-value for this test should also be 1.0. However, largely for readability, in the tables below these p-values are considered as missing (i.e., corresponding to a null test), denoted by a period ".". Note that StatXact adjusts for the variance, which would be 0. Then the significance levels of the test statistics are based on the result of a division by 0, i.e., undefined, and hence StatXact codes these p-values as missing.

For each species by gender by organ the number of animals analyzed and used in the statistical tests is presented first in the table. Note that indicating an organ was not examined requires a specification in the data. The tumor incidence for each organ is presented next, with the significance levels of the tests of trend, and the results of pairwise tests between the high, medium, (plus low-medium in mice) and low dose groups with controls. These statistical tests are conditioned on the animals actually evaluated, ignoring those not analyzed. When animals are selected for evaluation on the basis of criteris related to the endpoint, the assumptions for the computation of the p-values may not hold.

To adjust for the multiplicity of tests the so-called Haseman-Lin-Rahman rules discussed in Section 1.3.1.4 are often applied. That is, when testing for trend over dose and the difference between the highest dose group with a control group, to control the overall Type I error rate to roughly 10% for a standard two species, two sex study, one compares the unadjusted significance level of the trend test to 0.005 for common tumors and 0.025 for rare tumors, and the pairwise test to 0.01 for common tumors and 0.05 for rare tumors. Incidence in the control group is used to assess background tumor incidence, and thus whether a tumor is considered to be rare (background incidence <1%) or common. As also discussed in section 1.3.1.4, using these adjustments for other tests, like the pairwise tests for the differences between the control and the the low, low-medium (in mice), and medium dose groups can be expected to increase the overall type I error rate to some value above the nominal rough 10% level, possibly considerably higher than the nominal 10% rate.

Table A.2.1 in rats and Table A.2.2 in mice shows the tumors that had at least one mortality adjusted test whose nominal statistical significance was at least 0.05. Note that when one adjusts for multiplicity these nominally significant comparisons may not be statistically significant. Tables A.2.3 and A.2.4 display all incidences and statistical test results for male and

female rats, respectively, while Tables A.2.5 and A.2.6 present similar results in male and female mice. The p-values of the poly-k test are based on exact tests from StatXact as discussed above. As also noted above, the period '.' denotes the p-values of tests of dose groups with no tumors in any group.

Table A.2.1 Potentially Statistically Significant Neoplasms in Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
Male Rats								
PANCREAS								
ISLET CELL ADENOMA	7	17	14	12	.4375	.1540	.0698	.0169
Islet Adenoma/Carcinoma	9	19	18	14	.4572	.1710	.0352	.0220
Female Rats								
SUBCUTIS								
FIBROMA	0	1	4	1	.5547	.5000	.0256	.3333
Fibroma/Fibrosarcoma	0	1	5	1	.5727	.5000	.0070	.3333

Using the Haseman-Lin adjustment for multiplicity, in male rats the test of differences between the low dose and control groups in islet cell adenoma was not significant (i.e. $p = 0.0169 > 0.01$), nor were the tests between the low and medium dose groups in pooled islet cell adenoma/carcinoma (both $p = 0.0352, 0.0220 > 0.01$, respectively). From the incidence in the control group in female rats, we would define fibromas and pooled fibroma/fibrosarcomas as rare tumors. If one then accepts the increase in overall type I error to some value above the rough 10% level inherent in testing the medium and low dose groups, the pairwise tests between the medium dose group and controls in fibroma was statistically significant ($p = 0.0256 \leq 0.05$). Similarly, the pairwise test between the medium dose group and controls in pooled fibroma and fobrosarcoma would be statistically significant ($p = 0.007 \leq 0.05$). No other tests achieved the 0.05 level of significance.

In mice, results are similar. Again, those organ-tumor combinations with at least one nominally statistically significant result ($p \leq 0.05$) in mice are summarized below:

Table A.2.2 Potentially Statistically Significant Neoplasms in Mice

Organ	Incidence					Significance Levels				
Tumor	Cntrl	Low	LwMd	Med	Hi	Trend	HvsC	MvsC	LMvsC	LvsC
Male Mice										
TESTES										
# Evaluated	52	15	11	18	52					
INTERSTITIAL CELL TUMOUR	0	2	0	1	0	.7847	.	.	.2571	.0475
Female Mice										
LUNGS + BRONCHI										
# Evaluated	52	26	19	27	52					
PULMONARY ADENOMA	3	6	3	3	4	.7810	.5000	.1898	.3324	.0331
UTERUS										
# Evaluated	52	48	48	45	52					
Endo. Adenocarc./Strom. Sarcoma	0	0	0	1	2	.0338	.2476	.	.4639	.

Adjusting for multiplicity, with the inflated type I error, the tests of differences between the low dose and controls in interatitial cell tumor of the testes in male mice was barely statistically significant ($p = 0.0475 < 0.05$). while the test in pulmonary adenoma in female mice and was not significant ($p = 0.0331 > 0.01$), since it is not classified as a rare tumor. Further, the test of trend in pooled endometrial adenocarcinoma and stromal sarcoma was not statistically significant ($p = 0.0338 > 0.025$). No other tests in mice even achieved the nominal 0.05 level.

A complete table in male rats is presented below:

Table A.2.3 Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
ADRENALS								
# Evaluated	49	43	43	50				
CORTICAL CARCINOMA	1	0	0	1	.4686	.7576	1	1
MALIGNANT PHAEOCHROMOCYTOMA	3	2	1	1	.8560	.9437	.9242	.7760
PHAEOCHROMOCYTOMA	6	3	5	8	.1621	.4029	.6576	.8860
Pheochromocytoma, Any	7	5	6	8	.3312	.5171	.6329	.7526
BRAIN								
# Evaluated	50	28	35	50				
ASTROCYTOMA	0	0	1	0	.5215	.	.4118	.
MIXED GLIOMA	1	0	0	2	.2230	.5000	1	1
EYES								
# Evaluated	50	24	35	49				
MALIGNANT SCHWANNOMA	0	1	0	0	.6835	.	.	.3243
H-POIETIC TUMOUR								
# Evaluated	2	1	3	4				
HISTIOCYTIC SARCOMA	1	0	3	0	.9429	1	.4000	1
Hist. Sarcoma/Lymphoma[M]	2	0	3	3	.6000	1	.	1
LYMPHOID LEUKAEMIA	0	0	0	1	.4000	.6667	.	.
Leukemia	0	1	0	1	.4889	.6667	.	.3333
MALIGNANT LYMPHOMA	1	0	0	3	.1190	.6000	1	1
MYELOID LEUKAEMIA	0	1	0	0	.8000	.	.	.3333
HEAD								
# Evaluated	1	0	2	2				
SQUAMOUS CELL CARCINOMA	0	0	0	1	.4000	.6667	.	.
SQ.CELL CARC. OF ZYMBAL's GLAND	1	0	1	1	.9000	1	1	.
Squamous Cell Carc., Any	1	0	1	2	.6000	.	1	.
HEART								
# Evaluated	50	50	50	50				
MALIGNANT SCHWANNOMA	1	0	0	0	1	1	1	1
HINDLIMBS								
# Evaluated	3	2	3	2				
OSTEOSARCOMA	1	0	0	0	1	1	1	1
JEJUNUM								
# Evaluated	50	24	33	50				
ADENOCARCINOMA	1	0	0	0	1	1	1	1

Table A.2.3 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
KIDNEYS								
# Evaluated	50	50	50	50				
RENAL LIPOSARCOMA	1	0	0	0	1	1	1	1
RENAL MESENCHYMAL TUMOUR	1	0	1	0	.7513	1	.7525	1
Renal Lipoma/Liposarcoma	1	0	0	0	1	1	1	1
LIVER								
# Evaluated	50	47	47	50				
BENIGN LIVER CELL TUMOUR	0	1	0	2	.1125	.2475	.	.4845
Liver Cell Tumor [B]&[M]	0	1	3	3	.0842	.1212	.1100	.4845
MALIGNANT LIVER CELL TUMOUR	0	0	3	1	.2880	.5000	.1100	.
MAMMARY								
# Evaluated	50	28	36	50				
Fibroadenoma/Adenocarc.	3	0	0	2	.4971	.8189	1	1
MAMMARY ADENOCARCINOMA	2	0	0	2	.3564	.6913	1	1
MAMMARY FIBROADENOMA	1	0	0	0	1	1	1	1
PANCREAS								
# Evaluated	50	50	50	50				
EXOCRINE ADENOCARCINOMA	0	0	0	1	.2500	.5000	.	.
EXOCRINE ADENOMA	1	1	1	1	.5145	.7525	.7525	.7525
Exo. Adenoma/Adenocarc.	1	1	1	2	.2595	.5000	.7525	.7525
ISLET CELL ADENOMA	7	17	14	12	.4375	.1540	.0698	.0169
ISLET CELL CARCINOMA	4	3	4	2	.7809	.8978	.6425	.7820
Islet Adenoma/Carcinoma	9	19	18	14	.4572	.1710	.0352	.0220
PARATHYROIDS								
# Evaluated	42	47	42	43				
ADENOMA	0	0	1	0	.4885	.	.5000	.
PITUITARY								
# Evaluated	50	37	44	50				
Adenoma/Adenocarcinoma	25	24	25	20	.9636	.8862	.3252	.1222
PITUITARY ADENOCARCINOMA	1	1	2	0	.8139	1	.4517	.6725
PITUITARY ADENOMA	24	23	23	20	.9342	.8431	.4182	.1372
PROSTATE								
# Evaluated	50	25	34	49				
ADENOCARCINOMA	0	1	0	0	.6835	.	.	.3333
SKELETAL MUSCLE								
# Evaluated	2	2	2	2				
FIBROSARCOMA	0	1	1	0	.6429	.	.5000	.5000
SKIN								
# Evaluated	50	37	43	50				
FIBROMA	1	0	1	0	.7678	1	.7137	1
FIBROSARCOMA	0	0	1	0	.5167	.	.4624	.
Fibroma/Fibrosarcoma	1	0	2	0	.7208	1	.4430	1
INVERTED SQ. CELL PAPILLOMA	1	0	1	0	.7678	1	.7137	1
LIPOMA	1	1	0	0	.9240	1	1	.6725
SEBACEOUS ADENOMA	1	0	0	0	1	1	1	1
SEBACEOUS BASAL CELL CARCINOMA	1	0	0	0	1	1	1	1
SQUAMOUS CELL PAPILLOMA	2	2	0	2	.5196	.6913	1	.5707
Sq. Cell Papilloma	3	2	1	2	.6659	.8189	.9211	.7120

Table A.2.3 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Male Rats

Organ / Tumor	Incidence Cntrl	Low	Med	Hi	Significance Levels Trend	HvsC	MvsC	LvsC
SUBCUTIS								
# Evaluated	16	16	14	14				
FIBROMA	3	5	4	1	.9082	.9336	.4186	.3425
FIBROSARCOMA	1	0	0	0	1	1	1	1
Fibroma/Fibrosarcoma	4	5	4	1	.9425	.9693	.5743	.5000
LIPOMA	8	11	8	6	.8301	.7753	.4905	.2363
TESTES								
# Evaluated	50	50	50	50				
INTERSTITIAL CELL TUMOUR	3	4	1	4	.3566	.5000	.9413	.5000
THORAX								
# Evaluated	1	0	0	1				
FIBROMA	1	0	0	0	1	1	.	.
THYMUS								
# Evaluated	44	19	30	45				
THYMIC ADENOCARCINOMA	0	0	0	1	.3261	.5056	.	.
THYROIDS								
# Evaluated	50	50	50	50				
FOLLICULAR ADENOMA	6	0	1	0	.9985	1	.9938	1
FOLLICULAR CARCINOMA	0	0	1	0	.5000	.	.5000	.
Foll. Adenoma/-carc./Para.	16	6	8	14	.2479	.7435	.9831	.9965
Follicular Adenoma/-carcinoma	6	0	2	0	.9939	1	.9703	1
PARAFOLLICULAR CELL CARCINOMA	11	6	6	14	.0599	.3224	.9458	.9458
URINARY BLADDER								
# Evaluated	49	25	34	50				
PROSTATIC ADENOCARCINOMA	0	1	0	0	.6899	.	.	.3378
TRANSITIONAL CELL PAPILLOMA	0	0	1	0	.5316	.	.4096	.

A complete table in female rats is presented below:

Table A.2.4 Incidence and Significance Levels of all Tests on Neoplasms in Female Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
ADIPOSE TISSUE								
# Evaluated	4	4	1	4				
LIPOMA	0	1	0	0	.6923	.	.	.5000
ADRENALS								
# Evaluated	50	50	49	50				
MALIGNANT PHAEOCHROMOCYTOMA	0	0	0	1	.2513	.5000	.	.
PHAEOCHROMOCYTOMA	3	0	1	2	.4369	.8189	.9388	1
Pheochromocytoma, Any	3	0	1	3	.2264	.6611	.9388	1
BRAIN								
# Evaluated	50	38	28	50				
PINEAL BLASTOMA	1	0	0	0	1	1	1	1
H-POIETIC TUMOUR								
# Evaluated	1	0	0	3				
HISTIOCYTIC SARCOMA	1	0	0	2	1	1	.	.
LYMPHOID LEUKAEMIA	0	0	0	1	.7500	.7500	.	.
HEAD								
# Evaluated	1	2	4	0				
SEBACEOUS SQUAMOUS CELL CARCINO	0	0	1	0	.5714	.	.8000	.
SQUAMOUS CELL CARCINOMA	0	0	1	0	.5714	.	.8000	.
Squamous Cell Carc., Any	0	0	2	0	.2857	.	.6000	.
HINDLIMBS								
# Evaluated	1	1	1	2				
OSTEOSARCOMA	0	1	0	0	.8000	.	.	.5000
KIDNEYS								
# Evaluated	50	34	27	50				
RENAL LIPOMA	0	0	0	1	.3106	.5000	.	.
RENAL LIPOSARCOMA	0	1	0	0	.6894	.	.	.4048
Renal Lipoma/Liposarcoma	0	1	0	1	.3319	.5000	.	.4048
LIVER								
# Evaluated	50	47	49	50				
BENIGN LIVER CELL TUMOUR	0	0	0	1	.2551	.5000	.	.
MAMMARY								
# Evaluated	50	50	50	50				
Adenoma/Fibroad./Adenocarc.	33	32	34	32	.5777	.6623	.5000	.6623
MAMMARY ADENOCARCINOMA	9	13	7	5	.9597	.9261	.7930	.2348
MAMMARY ADENOMA	2	1	2	1	.6545	.8788	.6913	.8788
MAMMARY FIBROADENOMA	29	28	28	28	.5589	.6568	.6568	.6568
MAMMARY FIBROADENOMA W/ EPITH.	6	5	4	4	.7228	.8411	.8411	.7377
OVARIES								
# Evaluated	50	37	32	50				
THECAL CELL TUMOUR	0	1	0	0	.7041	.	.	.4253
PANCREAS								
# Evaluated	50	28	28	50				
ISLET CELL ADENOMA	5	0	4	3	.6368	.8657	.4113	1
ISLET CELL CARCINOMA	0	0	2	1	.2713	.5000	.1259	.
Islet Adenoma/Carcinoma	5	0	6	4	.5271	.7565	.1467	1
PAWS								
# Evaluated	11	7	7	2				
BASAL CELL CARCINOMA	0	0	1	0	.3333	.	.3889	.

Table A.2.4 (cont.) Incidence and Significance Levels of all Tests on Neoplasms in Female Rats

Organ	Incidence				Significance Levels			
Tumor	Cntrl	Low	Med	Hi	Trend	HvsC	MvsC	LvsC
PITUITARY								
# Evaluated	50	45	42	50				
Adenoma/Adenocarcinoma	34	34	29	28	.9669	.9256	.5477	.2791
PITUITARY ADENOCARCINOMA	2	4	3	4	.3275	.3389	.4171	.2897
PITUITARY ADENOMA	32	30	26	24	.9775	.9654	.6646	.4779
SKELETAL MUSCLE								
# Evaluated	0	1	1	0				
LIPOMA	0	0	1	0	.5000	.	.	.
STOMACH								
# Evaluated	50	34	27	50				
SQUAMOUS CELL PAPILLOMA-NONGL.	0	1	0	0	.6894	.	.	.4048
SUBCUTIS								
# Evaluated	8	4	7	8				
FIBROMA	0	1	4	1	.5547	.5000	.0256	.3333
FIBROSARCOMA	0	0	1	0	.5556	.	.4667	.
Fibroma/Fibrosarcoma	0	1	5	1	.5727	.5000	.0070	.3333
LIPOMA	4	3	1	2	.8995	.9406	.9814	.4242
TAIL								
# Evaluated	15	15	9	12				
FIBROSARCOMA	0	0	1	0	.4118	.	.3750	.
THYROIDS								
# Evaluated	50	50	50	50				
FOLLICULAR ADENOMA	0	1	0	1	.3128	.5000	.	.5000
FOLLICULAR CARCINOMA	1	0	0	0	1	1	1	1
Foll. Adenoma/-carc./Para.	10	11	9	11	.4234	.5000	.6945	.5000
Follicular Adenoma/-carcinoma	1	1	0	1	.5336	.7525	1	.7525
PARAFOLLICULAR CELL CARCINOMA	10	11	9	10	.5384	.5984	.6945	.5000
UTERUS								
# Evaluated	50	36	33	50				
ENDOMETRIAL STROMAL SARCOMA	0	1	0	0	.7041	.	.	.4186
SQUAMOUS CELL CARCINOMA	0	0	0	1	.2959	.5000	.	.
VAGINA								
# Evaluated	50	29	24	50				
FIBROMA	0	1	0	0	.6732	.	.	.3671

Table A.2.5 Incidence and Significance Levels of all Tests on Neoplasms in Male Mice

Organ	Incidence					Significance Levels				
Tumor	Cntrl	Low	LwMd	Med	Hi	Trend	HvsC	MvsC	LMvsC	LvsC
ADRENALS										
# Evaluated	52	14	12	22	50					
CORTICAL ADENOMA - TYPE B	1	1	0	0	0	.8813	1	1	1	.3818
H-POIETIC TUMOUR										
# Evaluated	1	0	1	0	1					
Lymphoma, Any	1	0	1	0	1	1	.	.	.	.
MALIGNANT LYMPHOMA	1	0	0	0	1	.6667	.	1	.	.
PLEOMORPHIC LYMPHOMA	0	0	1	0	0	.6667	.	.5000	.	.
KIDNEYS										
# Evaluated	52	52	50	52	52					
RENAL ADENOMA	1	1	0	0	0	.9600	1	1	1	.7524
RENAL CARCINOMA	1	0	0	0	0	1	1	1	1	1
Renal Adenoma/Carcinoma	1	1	0	0	0	.9600	1	1	1	.7524
LIVER										
# Evaluated	52	26	32	28	52					
BENIGN LIVER CELL TUMOUR	10	10	8	4	9	.8632	.6937	.3586	.8041	.0613
HAEMANGIOSARCOMA	0	0	1	0	0	.5895	.	.3810	.	.
Liver Cell Tumor [B]&[M]	12	11	11	4	11	.8816	.6814	.1901	.8930	.0691
MALIGNANT LIVER CELL TMR	2	2	3	0	2	.7145	.6912	.2801	1	.4074
LUNGS + BRONCHI										
# Evaluated	52	27	27	25	52					
PULMONARY ADENOCARCINOMA	4	4	2	0	1	.9757	.9717	.6743	1	.2674
PULMONARY ADENOMA	14	4	11	4	9	.8968	.9223	.1592	.9143	.9369
Pulmonary Adenoma/-carc.	16	7	12	4	10	.9551	.9440	.1691	.9556	.7591
PITUITARY										
# Evaluated	45	12	9	16	49					
ADENOMA	0	1	1	0	0	.7650	.	.1667	.	.2105
SKIN										
# Evaluated	52	13	9	17	52					
HAEMANGIOMA	0	0	0	1	0	.4825	.	.	.2464	.
PAPILLOMA	1	0	0	0	0	1	1	1	1	1
Systemic										
# Evaluated	52	52	52	52	52					
HAEMANGIOMA	0	0	0	1	0	.4000	.	.	.5000	.
HAEMANGIOSARCOMA	0	0	1	0	0	.6000	.	.5000	.	.
TESTES										
# Evaluated	52	15	11	18	52					
INTERSTITIAL CELL TUMOUR	0	2	0	1	0	.7847	.	.	.2571	.0475

Table A.2.6 Incidence and Significance Levels of all Tests on Neoplasms in Female Mice

Organ	Incidence					Significance Levels				
Tumor	Cntrl	Low	LwMd	Med	Hi	Trend	HvsC	MvsC	LMvsC	LvsC
H-POIETIC TUMOUR										
# Evaluated	5	3	4	5	5					
HISTIOCYTIC SARCOMA	0	2	2	2	2	.3283	.2222	.1667	.2222	.1071
LYMPHOID LEUKAEMIA	0	0	0	1	0	.4545	.	.	.5000	.
Lymphoma, Any	5	1	2	2	3	.6584	1	1	1	1
MALIGNANT LYMPHOMA	3	1	1	1	3	.2935	.7381	.9603	.9762	.9286
PLEOMORPHIC LYMPHOMA	2	0	1	1	0	.8942	1	.8810	.9167	1
HARDERIAN GLANDS										
# Evaluated	0	1	0	0	0					
ADENOMA	0	1	0	0	0	1	.	.	.	.
LIVER										
# Evaluated	52	16	16	25	52					
BENIGN LIVER CELL TUMOUR	0	0	1	1	1	.2974	.5000	.2353	.3247	.
Liver Cell Tumor [B]&[M]	1	0	1	1	2	.2819	.5000	.4179	.5468	1
MALIGNANT LIVER CELL TMR	1	0	0	0	1	.5430	.7524	1	1	1
LUNGS + BRONCHI										
# Evaluated	52	26	19	27	52					
PULMONARY ADENOCARCINOMA	2	1	1	2	0	.9109	1	.6133	.4227	.7095
PULMONARY ADENOMA	3	6	3	3	4	.7810	.5000	.1898	.3324	.0331
Pulmonary Adenoma/-carc.	5	7	4	5	4	.9242	.7561	.1865	.2171	.0509
MAMMARY										
# Evaluated	52	12	12	15	52					
MAMMARY ADENOCARCINOMA	1	0	0	0	0	1	1	1	1	1
OVARIES										
# Evaluated	52	40	38	46	52					
CYSTADENOMA	0	1	0	0	0	.7719	.	.	.	.4348
LUTEOMA	1	0	1	0	0	.8382	1	.6689	1	1
SKIN										
# Evaluated	52	13	14	16	52					
BASAL CELL TUMOUR	0	0	0	1	0	.4626	.	.	.2353	.
FIBROMA	0	1	0	0	0	.6463	.	.	.	.2000
UTERINE CERVIX										
# Evaluated	51	12	14	18	52					
FIBROMA	0	0	0	0	1	.3537	.5049	.	.	.
FIBROSARCOMA	0	0	1	0	0	.5714	.	.2154	.	.
Fibroma/Fibrosarcoma	0	0	1	0	1	.2786	.5049	.2154	.	.
UTERUS										
# Evaluated	52	48	48	45	52					
DECIDUOMA	0	1	0	0	0	.7878	.	.	.	.4800
ENDOMETRIAL ADENOCARCINOM	0	0	0	0	1	.2122	.5000	.	.	.
ENDOMETRIAL STROMAL SARCOMA	0	0	0	1	1	.1226	.5000	.	.4639	.
Endo.Adenocarc./Strom.Sarc.	0	0	0	1	2	.0338	.2476	.	.4639	.
FIBROMA	0	0	0	0	1	.2122	.5000	.	.	.
FIBROSARCOMA	0	1	0	0	0	.7878	.	.	.	.4800
LEIOMYOMA	1	0	0	0	0	1	1	1	1	1
VAGINA										
# Evaluated	52	12	13	17	52					
FIBROMA WITH ULCER	0	1	0	0	0	.6438	.	.	.	.1875

Appendix 3. References

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

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

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

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

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

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

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

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.

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 F THOMSON
01/23/2012

KARL K LIN
01/24/2012
Concur with review

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA Serial Number:	NDA 203-202 / SN 0000
Drug Name:	Droxidopa
Indication(s):	treatment of symptomatic neurogenic orthostatic hypotension (NOH) in patients with primary autonomic failure (Parkinson's Disease [PD], Multiple System Atrophy [MSA], and Pure Autonomic Failure [PAF]), Dopamine Beta-Hydroxylase (DBH) Deficiency, and Non-Diabetic Autonomic Neuropathy (NDAN)
Applicant:	Chelsea Therapeutics, Inc
Date(s):	Date of Document: September 28, 2011 PDUFA due date: March 28, 2011
Review Priority:	Priority
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	James Hung, Ph.D.
Medical Division:	Division of Cardiovascular and Renal Products, HFD-110
Clinical Team:	Melanie Blank, M.D.
Project Manager:	Anna Park, Pharm.D.

Keywords:

Analysis of Covariance, LOCF, Wilcoxon rank-sum test, Randomized withdrawal

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

The Chelsea-sponsored droxidopa clinical development program includes two randomized, placebo-controlled, double-blind studies (Studies 301 and 302). Only one study appears to demonstrate the efficacy of droxidopa.

Study 301 was an induction-design study with an initial open-label dose-titration period prior to a 7-day washout period, followed by a 7-day double-blind randomized treatment period in which patients were treated with either droxidopa or matching placebo. Despite a change in the primary endpoint when a majority of subjects (126 out of 162) completed the end of study visit, the results in Study 301 appeared to be consistent overall and showed the efficacy of droxidopa. Analyses on Orthostatic Hypotension Questionnaire (OHQ) composite score (primary endpoint), Orthostatic Hypotension Symptom Assessment (OHSA) item 1 score (original primary endpoint) and 3 minutes post-standing systolic blood pressure all showed statistical significance favoring droxidopa group.

Study 302 was a withdrawal-design study that included an initial open-label dose-titration period, a 7-day open-label treatment period, and a 14-day randomized-withdrawal period in which patients were treated with either droxidopa or matching placebo. The pre-specified primary endpoint, OHSA item 1 score, failed to show statistical significance with p-value of 0.51 indicating no signal. The standing blood pressure did not show that droxidopa was statistically better than placebo although there appeared to be a numeric trend favoring droxidopa. The post-hoc analysis using OHQ composite score seemed to suggest a signal favoring droxidopa. However, it remains a question whether OHQ composite score is a valid measurement for Neurogenic Orthostatic Hypotension (NOH) symptoms. Overall, study 302 did not provide support for the efficacy of droxidopa.

Additional study is needed to confirm the finding in Study 301.

2. INTRODUCTION

2.1 Overview

The Chelsea-sponsored droxidopa clinical development program includes two randomized, placebo-controlled, double-blind studies (Studies 301 and 302). Studies 301 and 302 were designed to enrich for responders in the study population by excluding patients during the open-label titration phase (prior to Randomization) who failed to demonstrate a symptomatic response in the OHSA Item 1 score and a physiological response in standing SBP.

Both studies enrolled patients with clinical diagnoses of symptomatic Neurogenic Orthostatic

Hypotension (NOH) associated with primary autonomic failure (PAF), Dopamine Beta Hydroxylase (DBH) Deficiency, or Non-Diabetic Neuropathy (NDAN) with a documented fall in SBP of at least 20 mmHg or in DBP of at least 10 mmHg within 3 minutes after standing.

Study 301 was a pivotal, Phase 3, multi-center, double-blind, randomized, placebo-controlled, parallel-group, induction-design study with an initial open-label dose-titration period (up to 14 days) prior to a 7-day washout period, followed by a 7-day double-blind randomized treatment period in which patients were treated with either their individually optimized dose of droxidopa or matching placebo. The primary efficacy variable for this study was changed from the original endpoint of Item 1 of OHSA to OHQ composite score. The change was reflected on the protocol dated December 2009.

Study 302 was a supportive, Phase 3, multi-center, double-blind, randomized, placebo-controlled, parallel-group, withdrawal-design study that included an initial open-label dose-titration period (up to 14 days), a 7-day open-label treatment period, and a 14-day randomized-withdrawal period in which patients were treated with either their individually optimized dose of droxidopa or matching placebo. Study 302 failed to meet its primary endpoint.

Table 1: List of phase 3 studies included in review

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
Study 301	Phase 3	7 days	7 days	80 in placebo and 82 in droxidopa	subjects with PAF, DBH Deficiency or and NDAN symptomatic NOH
Study 302	Phase 3	14 days	7 days	51 in placebo and 50 in droxidopa	subjects with PAF, DBH Deficiency or and NDAN symptomatic NOH

2.2 Data Sources

The sponsor's electronic data is stored under the directory \\Cdsesub5\evsprod\NDA203202\0000\m5\datasets. Specifically, data in Study 301 can be found under directory \\Cdsesub5\evsprod\NDA203202\0000\m5\datasets\noh301 and data in Study 302 can be found under \\Cdsesub5\evsprod\NDA203202\0000\m5\datasets\noh302.

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The reviewer was able to reproduce the primary analysis datasets from raw datasets. The sponsor submitted the SAS program code used to derive the primary analysis datasets and the reviewer is able to trace how the endpoint was derived from the original data source.

3.2 Evaluation of Efficacy

3.2.1 Study 301

3.2.1.1 Study Design and Endpoints

Study 301 was a pivotal, Phase 3, multi-center, double-blind, randomized, placebo-controlled, parallel-group, induction-design study with an initial open-label dose-titration period (up to 14 days) prior to a 7-day washout period, followed by a 7-day double-blind randomized treatment period in which patients were treated with either their individually optimized dose of droxidopa or matching placebo (Figure 1).

Following the screening period, eligible patients will then enter the open-label dose titration, where they will be treated with droxidopa and titrated to effect. Dose titration will begin at 100 mg TID of droxidopa and will be escalated in 100 mg TID increments until one or more of the following criteria are met:

1. The patient becomes asymptomatic (i.e. a score of “0” on item 1 of the OHSA) and has an improvement in standing systolic blood pressure (SBP) of at least 10 mmHg relative to baseline (all measurements made 3 minutes post standing);

2. The patient has a sustained SBP of greater than 180 mmHg or DBP of greater than 110 mmHg after 3 minutes of standing or after 5 minutes of sitting, OR a sustained SBP greater than 180 mmHg or DBP greater than 110 mmHg measured in the supine (head and torso elevated at approximately 30° from horizontal) position;

3. The patient is unable to tolerate side effects believed to be related to the study medication;

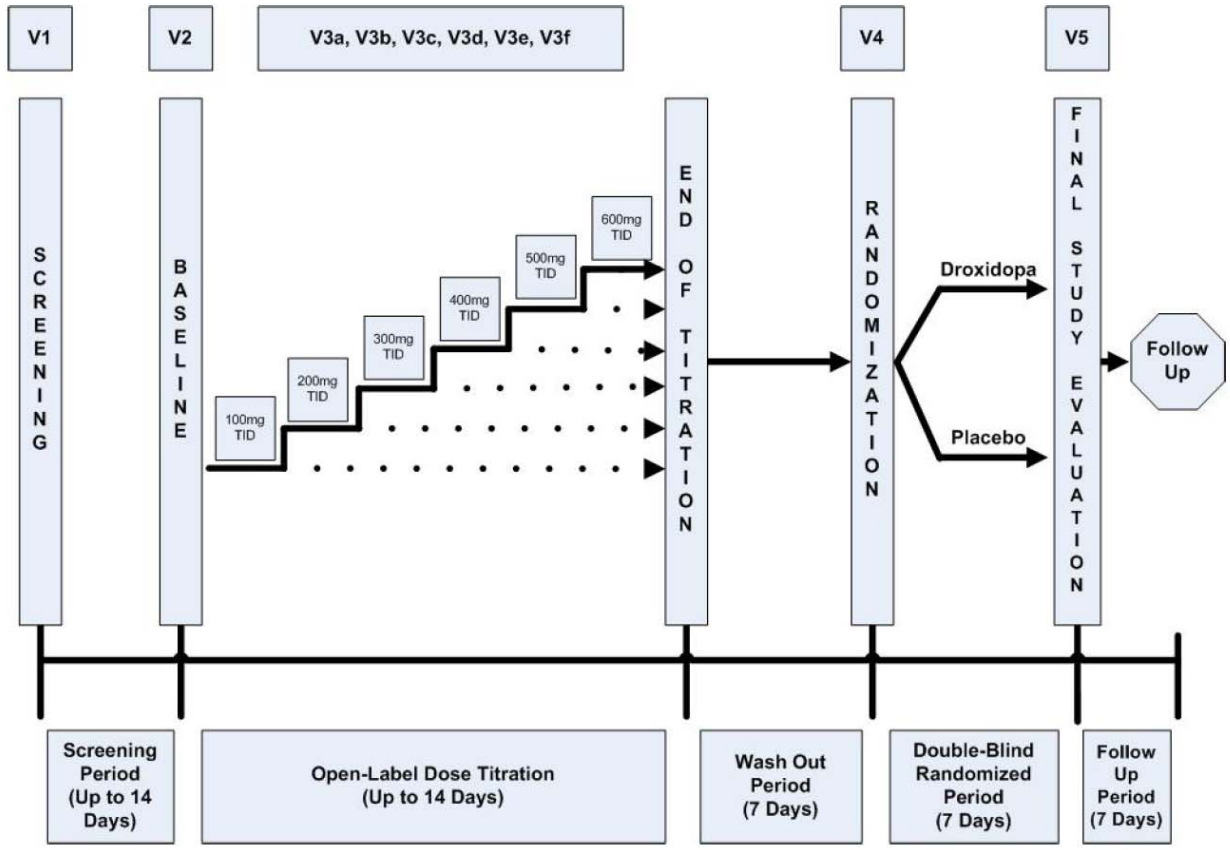
4. The patient reaches maximum dose of 600 mg TID (1800 mg/day) droxidopa.

Patients that meet any of the following criteria will be considered treatment failures and will not enter the treatment period of the study:

- Patients that meet criteria 2 or 3 and did not qualify as a responder at the previous lower dose;
- Patients that meet criteria 2 or 3 at the initial dose of 100 mg TID;
- Patients that meets criteria 4 and did not qualify as a responder at any dose.

Patients who were defined as being Responders at open-label titration period were entered into the washout period and randomized into the double-blind treatment period at the highest tolerated dose at which they qualified as a Responder.

Figure 1 Study Design (Study 301)



TID=Three times daily.

[Source: Figure 9-1 from Sponsor's clinical study report]

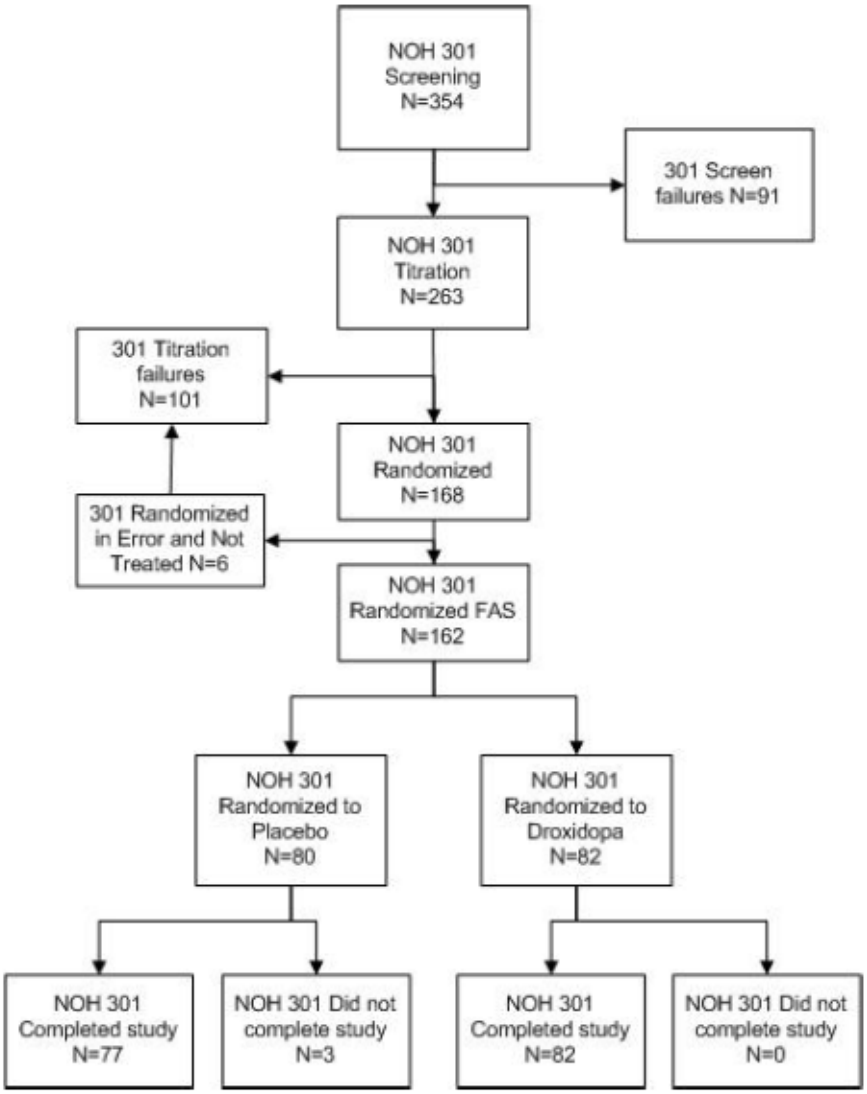
The primary efficacy variable was the mean change from Randomization to End of Study in the OHQ composite score. The OHQ composite is a global assessment of disease activity and is measured as the average of the OHSA composite and Orthostatic Hypotension Daily Activity Scale (OHDAS) composite scores. The OHSA scale measures symptoms associated with low blood pressure, using an 11-point scale (zero to 10), with more severe symptoms scoring higher. A zero score means that the symptom was not experienced. The scale assesses six symptoms: (1) Dizziness, (2) Problems with vision, (3) Weakness, (4) Fatigue, (5) Trouble concentrating, and (6) Head/neck discomfort. The OHDAS measures the impact of NOH symptoms on patients' ability to perform daily activities that require standing or walking. Patients were instructed to rate how their low blood pressure affected the daily activities including: (1) Standing for a short time, (2) Standing for a long time, (3) Walking for a short time, and (4) Walking for a long time.

3.2.1.2 Patient Disposition, Demographic and Baseline Characteristics

A total of 263 patients were titrated to determine Responders, and 168 patients were randomized. Of the patients randomized, 6 patients who did not receive blinded study drug were excluded from the FAS. The FAS here is in fact a mITT population that included patients who were randomized and received at least 1 dose of double-blind study drug. The FAS consisted of 162 patients: 80 patients randomized to placebo and 82 patients randomized to droxidopa.

Of the 101 patients who did not receive double-blind study drug, the most common reason for discontinuation as determined by the Investigator was treatment failure (50 [49.5%] patients).

Figure 2 Patient Distribution and Study Population (Study 302)



[Source: Figure 10-1 from Sponsor's clinical study report]

Demographic characteristics were overall similar between the placebo and droxidopa groups. The mean ages were 55.8 and 57.3 years for patients in the placebo and droxidopa groups, respectively. Patients only treated in the open-label titration phase (N=101) was older (64.6 years). There was a nearly equal distribution of male and female patients in the placebo and droxidopa groups, but a greater proportion of males in the open-label phase. The patients were predominantly white in the study. The droxidopa group was mostly composed of patients with a primary diagnosis of PD (43.2%), PAF (32.1%), or MSA (17.3%). Similar proportions were observed in the placebo group (38.3%, 34.6%, and 14.8%, respectively). Approximately 60% of patients were enrolled at non-US sites. For patients only treated in the open-label phase, 52.5% were treated outside the US.

Table 2 Patient Demographics (Study 301)

		Double-Blind Phase	
	Open-Label Phase[1] (N=101)	Placebo (N=81)	Droxidopa (N=81)
Sex [n (%)]			
Male	64 (63.4)	43 (53.1)	41 (50.6)
Female	37 (36.6)	38 (46.9)	40 (49.4)
Race [n (%)]			
White	99 (98.0)	76 (93.8)	81 (100.0)
Black/African American	2 (2.0)	1 (1.2)	0
Asian	0	1 (1.2)	0
Hispanic/Latino	0	3 (3.7)	0
Primary Clinical Diagnosis [n (%)]			
Parkinson's Disease	45 (44.6)	31 (38.3)	35 (43.2)
Multiple System Atrophy	18 (17.8)	12 (14.8)	14 (17.3)
Pure Autonomic Failure	33 (32.7)	28 (34.6)	26 (32.1)
Dopamine Beta Hydroxylase Deficiency	0	0	0
Non-Diabetic Autonomic Neuropathy	2 (2.0)	6 (7.4)	2 (2.5)
Other Diagnosis	3 (3.0)	4 (4.9)	4 (4.9)
Age (Years) at Screening			
Mean (SD)	64.6 (15.40)	55.8 (19.94)	57.3 (16.98)
Min, Max	19, 91	18, 87	20, 84
Weight (kg)			
N	99	80	80
Mean (SD)	74.85 (15.062)	74.96 (14.413)	74.10 (14.516)
Min, Max	44.0, 113.4	47.0, 110.0	46.0, 103.0
Geographic Region [n (%)]			
US	48 (47.5)	33 (40.7)	32 (39.5)
Non-US	53 (52.5)	48 (59.3)	49 (60.5)

1. Patients who were titrated in the open-label phase but not randomized are included only in the open-label droxidopa column. This also includes 6 patients who received study treatment during the open-label titration phase and who were randomized but never received double-blind drug

* Statistics on droxidopa and placebo groups are based on actual treatment received

[Source: Table 11-2 in sponsor's clinical study report, verified by the reviewer]

The mean Baseline OHQ composite scores were 5.62 and 5.96 units for the placebo and droxidopa groups, respectively. The mean SBP post-standing at 3 minutes was 90.7 and 90.8 mmHg for the placebo and droxidopa groups, respectively.

Table 3 Patient Baseline Characteristics (Study 301)

Parameter	Placebo (N=80)	Droxidopa (N=82)
Baseline OHQ Composite Score		
n	79	81
Mean (SD)	5.62 (1.98)	5.96 (1.67)
Min, Max	1.2, 9.8	2.0, 9.6
Baseline SBP upon Standing +3 Minutes (mmHg)		
n	80	82
Mean (SD)	90.7 (16.83)	90.8 (15.63)
Min, Max	50, 130	45, 142

[Source: Table 11-3 in sponsor's clinical study report, verified by the reviewer]

3.2.1.3 Statistical Methodologies

Primary efficacy endpoint is the mean change in the OHQ composite score. The OHQ composite score is computed as the average of the OHSA composite and OHDAS composite scores from the OHQ. At a given time point, the OHSA composite is the average of the symptom scores at that time for those symptoms present at baseline (e.g., if five symptoms were marked as present at baseline (i.e. score >0) then, the OHSA composite is the sum of the scores of those symptoms at the specified time point divided by 5). The OHDAS composite is the average of the activities that are scored at the same time point. Activities marked as zero or 'cannot be done for other reasons' at baseline are not included in the analysis. Where patients have a score for an OHDAS activity at baseline (i.e. score >0), OHDAS activities marked as 'cannot be done for other reasons' or without a value at randomization or end of study visits are imputed using LOCF. Each of the OHSA and OHDAS items are evaluated on a scale of 0 to 10. Thus the OHQ composite is a score that ranges from 0 to 10.

In the primary analysis, the droxidopa and placebo groups were compared using an analysis of covariance (ANCOVA). The missing data were imputed using LOCF. Since there is only one assessment of the OHQ following randomization, patients who have a missing value at post-randomization day 7 will be assumed to have a change from randomization equal to 0.

According to the SAP, the hierarchy of endpoints was defined as follows:

Primary efficacy endpoint is the mean change in the OHQ composite score (using LOCF and the FAS) from Randomization to End of Study (p=0.003).

Secondary efficacy endpoints include follows:

1. The change in OHDAS composite score for Items 1-4 (calculated as the arithmetic average of Items 1-4) from Randomization to End of Study;
2. The change in OHSA composite score for Items 1-6 (calculated as the arithmetic average of Items 1-6 with a Baseline score greater than 0) from Randomization to End of Study;
3. The change in OHDAS Item 1 (standing short time) from Randomization to End of Study;
4. The change in OHDAS Item 3 (walking short time) from Randomization to End of Study;
5. The change in OHSA Item 1 (dizziness) from Randomization to End of Study;
6. Improvement in the End of Study scores for the patient-rated CGI-S.

Due to the fact that the analysis performed for patient-rated CGI-S was not significant, no formal statistical testing was performed for subsequent efficacy endpoints.

The sponsor had a meeting with the Division on November 18, 2009 to discuss the change on primary endpoint. The primary efficacy variable for this study was changed from the original endpoint of Item 1 of the OHSA to the OHQ composite score. The change was reflected on the protocol dated December 2009. According to the sponsor, "Item 1 (dizziness) of the OHSA as a primary endpoint was insensitive to determining treatment effects in Study 302." Also "the performance of the OHQ composite was shown in a *post hoc* analysis to be superior to the performance of the OHSA Item 1 (dizziness) with regard to determining a treatment effect" in Study 302.

The sample size was also increased (from 118 patients to 150 patients) at the time of the change in the primary endpoint to power the study appropriately based on the new primary endpoint and the treatment effects reported with Study 302.

3.2.1.4 Results and Conclusion

The mean change in the OHQ composite score from Randomization to End of Study showed statistical significance favoring droxidopa (p=0.003). At End of Study, the mean decrease in OHQ composite score in droxidopa patients was 0.9 more than the mean decrease in OHQ score in placebo (Table 4).

Table 4 Summary of OHQ Composite Score (Full Analysis Set with LOCF)

	Placebo (N=80)	Droxidopa (N=82)	ANCOVA[3]
Randomization (Visit 4)			
N[4]	79	81	
Mean (SD)	4.97 (2.41)	5.11 (1.96)	
Min, Max	0.7, 9.8	0.9, 9.1	
End of Study (Visit 5)			
N	79	81	
Mean (SD)	4.04 (2.61)	3.29 (2.20)	
Min, Max	0.0, 9.8	0.0, 8.4	
Change from Randomization to End of Study			
N	79	81	0.003
Mean (SD)	-0.93 (1.69)	-1.83 (2.07)	
Min, Max	-7.5, 2.6	-6.2, 4.4	

[Source: Sponsor's clinical study report Table 11-5, verified by the reviewer]

In order to assess the impact of missing data on the primary analysis, the primary efficacy analysis was repeated excluding patients who had missing data for the primary endpoint.

Table 5 Summary of OHQ Composite Score (Full Analysis Set with Missing Data Excluded)

	Placebo (N=80)	Droxidopa (N=82)	ANCOVA[2]
Randomization (Visit 4)			
N[3]	79	80	
Mean (SD)	4.96 (2.43)	5.07 (1.92)	
Min, Max	0.1, 9.8	0.9, 9.1	
End Of Study (Visit 5)			
N	79	81	
Mean (SD)	4.05 (2.61)	3.29 (2.20)	
Min, Max	0.0, 9.8	0.0, 8.4	
Change from Randomization to End of Study			
N	79	80	0.002
Mean (SD)	-0.92 (1.73)	-1.84 (2.08)	
Min, Max	-7.5, 3.5	-6.2, 4.4	

[Source: Sponsor's clinical study report Table 11-6, verified by the reviewer]

Similar analysis was done for Per Protocol population, in which subjects with sufficiently serious violations/deviations were excluded, and result was consistent with significant p-value favoring droxidopa.

The reviewer did a sensitivity analysis using subjects who had End of Study visit before the decision of changing the primary endpoint. The last patient who completed before primary efficacy endpoint change had End of Study visit on September 28, 2009. A total of 126 subjects

completed End of Study visit before the date (64 in droxidopa and 62 in placebo). Two subjects had missing values on OHQ score. Similar analysis as the primary analysis was performed based on the 124 observations. The mean difference in the change of OHQ composite score between two groups is 0.8 and the result is nominally significant (p=0.02).

The original proposed primary endpoint, the change in OHSA item 1 score, also showed consistent results. The change in OHSA Item 1 (dizziness) from Randomization to End of Study had p-value of 0.001 using Wilcoxon rank sum test. Even if the subjects who completed End of Study visit before primary efficacy endpoint change were used (126 subjects), the conclusion remained the same (p=0.01).

Table 6 Summary of OHSA item 1 score

	Placebo			Droxidopa		
	N	Mean	STD	N	Mean	STD
Randomization	80	5.43	2.9	82	5.39	2.5
End of Study	80	4.30	3.1	82	2.98	2.7
Change from randomization to End of Study	80	-1.13	2.6	82	-2.41	3.2

To further assess the impact of excluding items with score of 0 at baseline, sensitivity analysis was also performed on OHQ composite score averaging all values reported by patients at Baseline and endpoint (including values for those symptoms that were 0 at any time during the trial). The p-value from ANCOVA model is 0.003.

Patients receiving droxidopa experienced a significant change from Randomization to End of Study in 3 minutes post-standing SBP values during the Orthostatic Standing Test (OST) compared with placebo. The mean change in standing SBP was 11.2 mmHg (SD=22.89 mmHg) following treatment with droxidopa and 3.9 mmHg (SD=16.28 mmHg) following treatment with placebo. The p-value was 0.02 based on ANCOVA model. The sponsor reported a p-value less than 0.001, which was based on a rank statistics using non-parametric ANCOVA. The reviewer found the sponsor's analysis unjustified since ANCOVA model is robust enough and should be appropriate for the dataset. Nevertheless, the change from randomization in SBP 3 minutes post standing appears to be significant and the conclusion remains the same.

Patients receiving droxidopa did not experience significant improvements from Randomization to End of Study in 3 minutes post-standing DBP compared with placebo. There was a numeric trend favoring the droxidopa group. The mean change in DBP is 5.5 mmHg (SD=13.4 mmHg) following treatment with droxidopa and 3.4 mmHg (SD=10.4 mmHg) following treatment with placebo.

There were two global assessments measured in the trial. One is CGI-I and the other is CGI-S. CGI-I is a 7-point scale ranging from a score of 1 (very much improved) to 7 (very much worse), with no change in the middle, and assesses the improvement in relation to the baseline evaluation. CGI-S is a 7-point scale ranging from a score of 1 (no symptoms) to 7 (severe symptoms). A reduction in score over a period of time would be considered an improvement in symptoms. Neither global assessment showed statistical significance for droxidopa over placebo.

In this reviewer's view, CGI-I questionnaire in the trial can be confusing to patients. Specifically, in both Randomization visit and End of Study visit, the questionnaire asked subjects to compare with their conditions at baseline visit (Visit 2), not the previous visit. After several visits in the open-label titration period and a 7-day washout period, subjects may not be able to recall their exact conditions several visits away, especially when the questionnaire was given at End of Study visit. This may be a partial reason why CGI-I did not show a statistical significant finding. The reviewer was not able to explain why CGI-S did not show any statistical significant finding. Droxidopa group did perform numerically better than the placebo group in both CGI-I and CGI-S assessments.

Analyses on OHQ composite score (primary endpoint), OHSA item 1 score (original primary endpoint) and 3 minutes post-standing SBP (clinical endpoint) and several sensitivity analyses on OHQ score as well as OHSA item 1 score all showed statistical significance favoring the droxidopa group. Although analyses on DBP, CGI-S, and CGI-I did not demonstrate statistical significance, there was a numeric trend favoring droxidopa. Overall, the results in Study 301 appear to be consistent.

3.2.2 Study 302

3.2.2.1 Study Design and Endpoints

Study 302 was a supportive, Phase 3, multi-center, double-blind, randomized, placebo-controlled, parallel-group, withdrawal-design study that included an initial open-label dose-titration period (up to 14 days), a 7-day open-label treatment period, and a 14-day randomized-withdrawal period in which patients were treated with either their individually optimized dose of droxidopa or matching placebo (**Figure 3**).

Eligible patients then entered the open-label dose-titration, where they were treated with droxidopa and titrated to effect. Dose titration began at 100 mg three times daily (TID) of droxidopa and was escalated in 100 mg TID increments until one or more of the following criteria (stopping rules) were met:

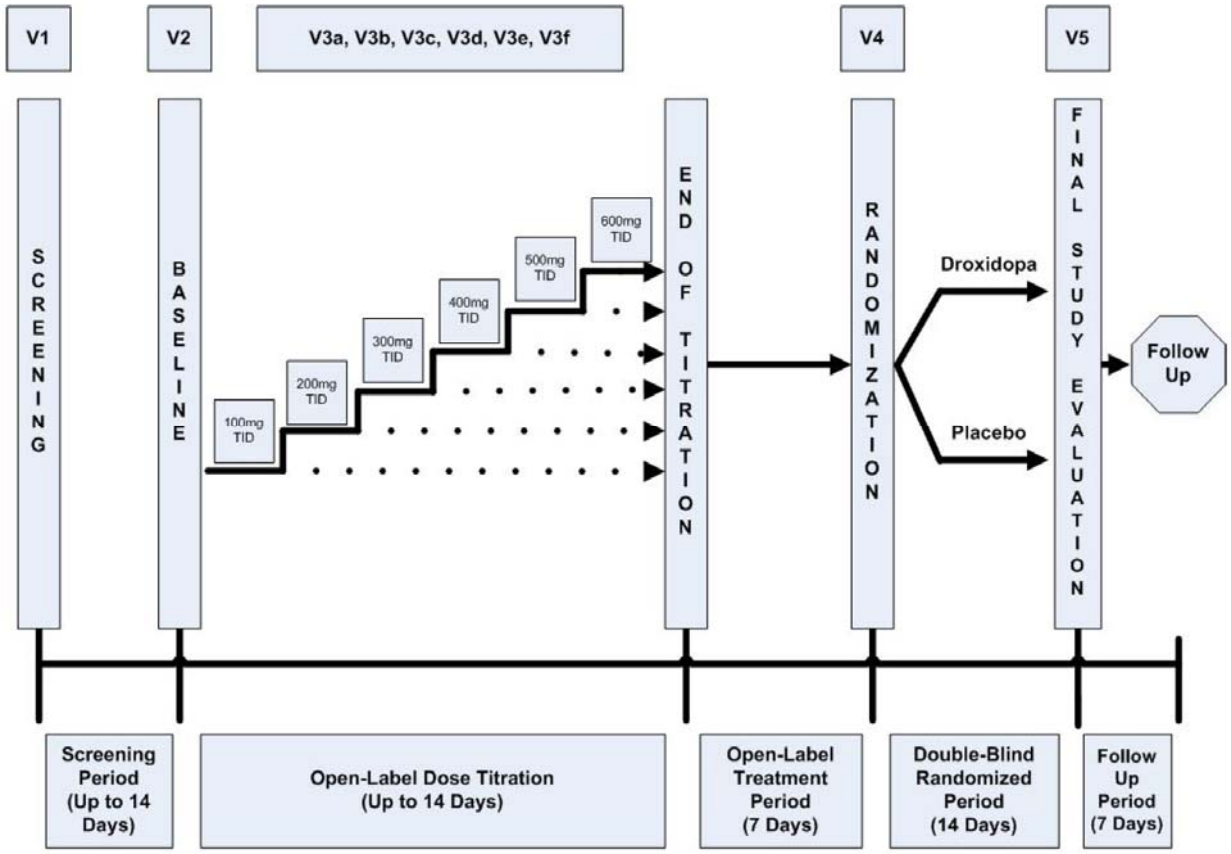
1. The patient became both asymptomatic (i.e., a score of "0" on Item 1 [dizziness] of the OHSA) and had an improvement in standing SBP of at least 10 mmHg relative to Baseline (all measurements made 3 minutes post-standing);
2. The patient had a sustained SBP of greater than 180 mmHg or DBP of greater than 110 mmHg after 3 minutes of standing or after 5 minutes of sitting, OR a sustained SBP greater than 180 mmHg or DBP greater than 110 mmHg measured in the supine (head and torso elevated at approximately 30° from horizontal) position;
3. The patient was unable to tolerate side effects believed to be related to the study drug;
4. The patient reached the maximum dose of 600 mg TID (1800 mg/day) droxidopa.

In order to accommodate the dose titration, the study included additional visits (Visits 3a, 3b, 3c, etc.) which may have been a single visit or as many as 6 visits. Patients were encouraged to schedule these visits on a daily basis; however, they were allowed to complete them over a 2-week period. At each visit, patients were required to undergo an orthostatic standing test (OST) to be conducted 3 hours after their morning dose of study treatment and complete Item 1 (dizziness) of the OHSA.

Following the open-label dose-titration, patients who demonstrated a symptomatic and BP response entered into a 7-day open-label treatment period at their titrated dose of droxidopa.

Following the open-label treatment period, patients who continued to show a symptomatic benefit entered into a 14-day double-blind treatment period and were randomized to treatment with either their titrated dose of droxidopa or matching placebo.

Figure 3 Study Design (Study 302)



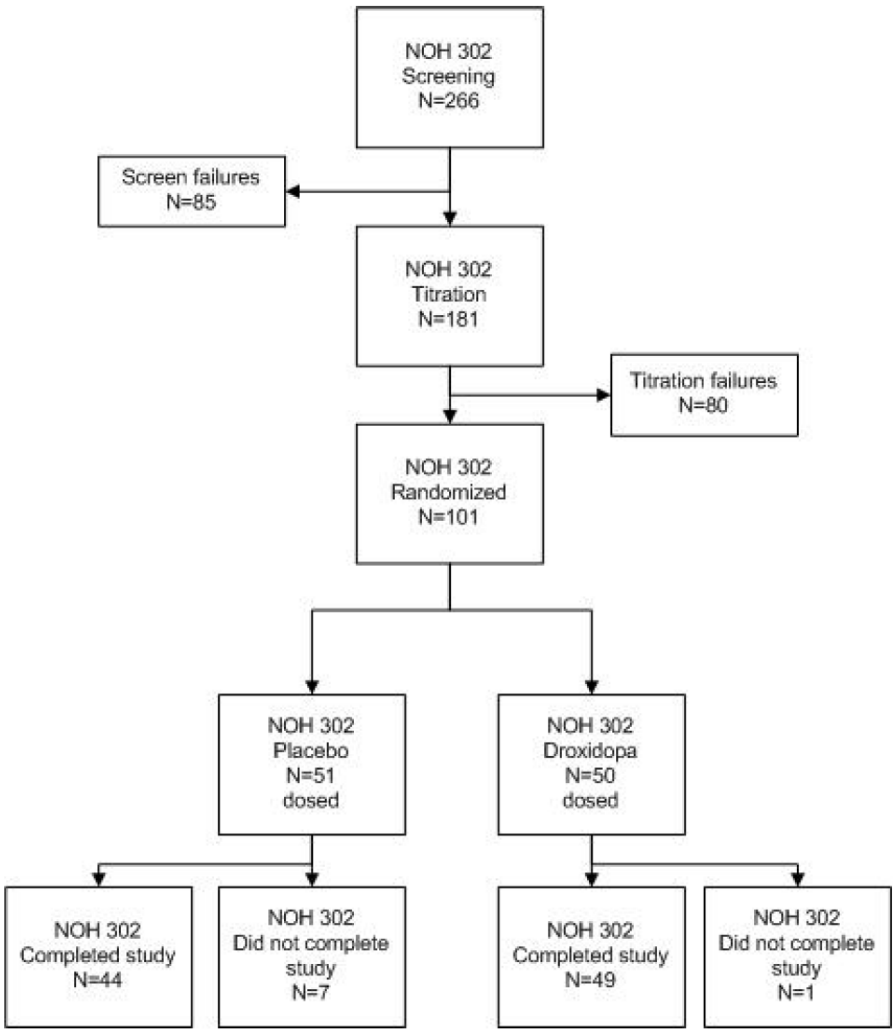
[Source: Figure 9-1 in the sponsor's clinical study report for Study 302]

The primary efficacy variable was the mean change from Randomization to End of Study in the OHSA Item 1 (dizziness) Score.

3.2.2.2 Patient Disposition, Demographic and Baseline Characteristics

A total of 181 patients received at least one dose of the study drug and were included in the Safety Set. Of these 181 patients, 101 patients were randomized (placebo: 51 patients; droxidopa: 50 patients) and comprise the FAS.

Figure 4 Patient Disposition (Study 302)



[Source: Sponsor's clinical study report for Study 302]

Of the 80 patients treated with droxidopa during the open-label phase who were not randomized, 55 subjects discontinued due to treatment failure

Of the 80 patients treated with droxidopa during the open-label phase who were not randomized, the most common reason for discontinuation was treatment failure (55 patients [68.8%]), followed by AEs (13 patients [16.3%]), protocol violations (4 patients [5.0%]), and withdrawal of consent (4 patients [5.0%])

Of the 101 patients who were randomized and received blinded study drug, 14 did not complete the study per protocol as defined by the site: 8 (15.7%) placebo-treated patients and 6 (12.0%) droxidopa-treated patient).

Table 7 Demographics Characteristics (Study 302)

		Double-Blind Phase	
	Open-Label Phase [2] (N=80)	Placebo (N=51)	Droxidopa (N=50)
Primary Clinical Diagnosis [n (%)]			
PD	38 (47.5)	23 (45.1)	21 (42.0)
MSA	21 (26.3)	13 (25.5)	17 (34.0)
PAF	18 (22.5)	10 (19.6)	8 (16.0)
DBH Deficiency	0	1 (2.0)	0
Non-Diabetic Neuropathy	2 (2.5)	3 (5.9)	2 (4.0)
Other Diagnosis	1 (1.3)	1 (2.0)	2 (4.0)
Age (Years) at Screening			
Mean (SD)	69.5 (9.74)	66.6 (11.25)	63.1 (13.76)
Min, Max	37, 86	40, 88	24, 88
Gender [n (%)]			
Male	45 (56.3)	32 (62.7)	30 (60.0)
Female	35 (43.8)	19 (37.3)	20 (40.0)
Region [n (%)]			
US	53 (66.3)	32 (62.7)	25 (50.0)
Non-US	27 (33.8)	19 (37.3)	25 (50.0)
Race [n (%)]			
White	79 (98.8)	48 (94.1)	49 (98.0)
Asian	0	1 (2.0)	1 (2.0)
American Indian/Alaskan Native	0	1 (2.0)	0
Hispanic/Latino	1 (1.3)	1 (2.0)	0
Weight (kg)			
N	79	50	50
Mean (SD)	75.71 (17.86)	73.02 (14.24)	76.66 (20.29)
Min, Max	45.4, 177.8	38.6, 99.0	47.0, 183.0

[Source: Table 11-2 in Sponsor's clinical study report for Study 302, verified by the reviewer]

The mean ages were 63.1 and 66.6 years for patients in the droxidopa and placebo groups, respectively. The majority of patients were male (60.0% and 62.7% for the droxidopa and placebo groups, respectively) and the patients were predominantly White. The demographic characteristics of the patients who received open-label treatment but were not randomized were older (69.5 years) than those of the randomized population.

The mean Baseline OHSA item 1 scores were 6.3 and 6.6 for the placebo and droxidopa groups, respectively. The mean SBP post-standing at 3 minutes was 88 and 87 mmHg for the placebo and droxidopa groups, respectively.

Table 8 Patient baseline Disease Severity (Study 302)

Parameter	Placebo (N=51)	Droxidopa (N=50)
Baseline OHSA Item 1 Score		
n	51	50
Mean (SD)	6.3 (2.27)	6.6 (2.01)
Min, Max	2, 10	3, 10
Baseline SBP upon Standing +3 Minutes (mmHg)		
n	50	50
Mean (SD)	88.0 (19.04)	87.0 (17.60)
Min, Max	50, 130	37, 116

[Source: Table 11-3 in Sponsor's clinical study report for Study 302, verified by the reviewer]

3.2.2.3 Statistical Methodologies

For the primary analysis, the droxidopa and placebo groups are compared using the Wilcoxon rank-sum test. The Full Analysis Set will be used for the primary analysis. Missing data were imputed using the LOCF method. Since there was only one assessment of the OHSA following Randomization, patients who had a missing value at post-Randomization Day 14 were assumed to have had a change from Randomization equal to 0. In order to assess the impact of missing data on the primary analysis, the primary efficacy analysis will be repeated excluding patients who have missing data for the primary endpoint.

The study was conducted at 71 centers. Patients from all centers were pooled and no adjustment or stratification for center was used.

The change from baseline in SBP and DBP will in the droxidopa and placebo groups were compared using Wilcoxon rank-sum tests.

In order to control the overall type I error, statistical significance of the primary and key secondary efficacy endpoints were evaluated using a hierarchical (or "gatekeeping") testing. Because the primary endpoint was not statistically significant, the hierarchical procedure was not utilized. The testing order was as follows:

1. Primary efficacy endpoint: change from randomization to end of study in the score of OHSA Item 1 (using LOCF and the Full Analysis Set).

2. Secondary efficacy endpoints:

- The change from randomization to the end of randomized treatment in the scores of OHSA and OHDAS:
    - OHSA item 4 (fatigue)
    - OHSA item 3 (weakness)
    - OHSA item 2 (vision).
    - OHSA item 5 (concentration).
    - OHSA item 6 (head/neck discomfort).
    - OHDAS composite score for items 1-4 (calculated as the arithmetic average of items 1-4).
    - OHSA composite score for items 1-6 (calculated as the arithmetic average of items 1-6).
    - OHSA composite score for items 2-6 (calculated as the arithmetic average of items 2-6).

- Change from randomization to end of treatment in SBP decrease upon orthostatic challenge.

3.2.2.4 Results and Conclusion

The mean change in the OHSA Item 1 Score from Randomization to End of Study was not significantly different between the droxidopa and placebo treatment groups (p=0.509). The pre-specified sequential testing would stop at this point since the primary endpoint was not significant. The sponsor attributed the failure to the relatively short length of the withdrawal phase. They argued that the withdrawal period may be insufficient to eliminate replenished Norepinephrine levels as suggested by the Norepinephrine pharmacokinetic data from placebo-treated patients. According to the sponsor, the continued clinical effects may have confounded the assessment of efficacy in this study. 14-day randomized withdrawal period may seem short. On the other hand, Study 301 had only 7-day washout period and a 7-day double-blind randomized treatment period. The washout period and treatment period were both short. The total length of the two periods is also 14 days. The short washout and treatment period, however, did not seem to hinder Study 301 from showing the efficacy of droxidopa.

From Randomization visit to End of Study visit, patients receiving droxidopa experienced a mean change in standing SBP of -7.6 mmHg during the OST versus a -5.2 mmHg change for patients receiving placebo (p=0.680). Interestingly, this is counter intuitive since this is a randomized withdrawal study and placebo treated patients were expected to have a decreased blood pressure more than the patients who were still on treatment. There was no difference in change in standing DBP between the two treatment groups.

Effective 26 February 2009, Study 302 was resized from 118 to 82 total patients. The initial sample size calculation for Study 302 estimated a standard deviation of 3.0 for the primary endpoint (i.e., OHSA Item 1). Subsequent data from other studies enabled a re-evaluation of the standard deviation, which resulted in lowering the estimate from 3.0 to 2.5. Using an overall 0.05 two-sided significance level, a new sample size of 41 evaluable patients in each randomized

treatment group in a 1:1 ratio (i.e., 82 patients in total) was determined to have 80% power to detect a difference of 1.6 points between placebo- and droxidopa-treated patients with respect to change from Randomization to End of Study in OHSA Item 1.

The sponsor subsequently did a post-hoc analysis using the OHQ composite score as an additional assessment of the effect of droxidopa. A p-value of 0.013 was reported using the Wilcoxon rank-sum test. Droxidopa-treated patients had a mean change of 0.11 units in their OHQ composite score compared with a change of 1.22 units in placebo-treated patients, resulting in a difference of 1.11 units favoring droxidopa. According to the sponsor, this result led to the primary endpoint change in Study 301. Based on the reviewer's calculation using ANCOVA model, which is the same analysis used for OHQ score in Study 301, the p-value is 0.04.

The significant result of the OHQ composite score appeared to be driven by the OHDAS subscale. The OHDAS composite alone showed statistical significance favoring droxidopa by ANCOVA model in Study 302. There are questions on whether the OHDAS comprehensively measures the impact of NOH symptoms on patients' daily activities. Please refer to the SEALD review for further details. It remains a question whether OHQ composite score is a valid measurement for NOH symptoms. Also it is unknown how much difference in OHQ score should we observe between the two treatment groups to reflect real clinical benefit.

Study 302 failed to show the efficacy of droxidopa. The pre-specified primary endpoint, OHSA item 1 score, failed to show any statistical significance with p-value = 0.68 showing no signal. The standing blood pressure did not show that droxidopa is statistically better than placebo although there was a numeric trend favoring droxidopa.

3.3 Evaluation of Safety

Please refer to the clinical review for safety evaluation.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

Since subjects in both trials were predominantly white (>95%), subgroup analysis on race is not performed in both trials.

In Study 301, droxidopa does not appear to have an effect on subjects over 75 years old. The female subjects seem to have less treatment effect numerically. The non US region seems to have a slightly larger treatment effect compared with US.

In Study 302, less treatment effect of droxidopa in subjects over 75 years old and female subjects can also be observed.

The reviewer did not perform formal statistical testing on the subgroups due to the limit of sample size. The numeric differences among subgroups could be due to chance.

Table 9 Subgroup analysis on OHQ composite score by age in Study 301

	Droxidopa			Placebo		
Age	N	Mean	STD	N	Mean	STD
Age<=65	53	-2.08	1.82	47	-0.91	1.69
65<Age<=75	17	-1.57	3.02	19	-0.69	1.55
Age>75	11	-1.03	1.11	13	-1.39	1.94

Table 10 Subgroup analysis on OHQ composite score by gender in Study 301

	Droxidopa			Placebo		
Gender	N	Mean	STD	N	Mean	STD
Male	41	-2.13	1.98	41	-0.83	1.41
Female	40	-1.52	2.13	38	-1.04	1.97

Table 11 Subgroup analysis on OHQ composite score by region in Study 301

	Droxidopa			Placebo		
Region	N	Mean	STD	N	Mean	STD
US	32	-1.54	2.33	31	-0.83	2.14
Non US	49	-2.02	1.87	48	-1.00	1.35

Table 12 Subgroup analysis on OHQ composite score by age in Study 302

	Droxidopa			Placebo		
Age	N	Mean	STD	N	Mean	STD
Age<=65	25	0.06	1.79	26	1.41	2.75
65<Age<=75	17	0.17	2.49	13	0.89	2.41
Age>75	8	0.56	2.59	12	0.87	1.47

Table 13 Subgroup analysis on OHQ composite score by gender in Study 302

	Droxidopa			Placebo		
Gender	N	Mean	STD	N	Mean	STD
Male	30	-0.14	1.52	32	1.03	2.43
Female	20	0.66	2.81	19	1.36	2.38

Table 14 Subgroup analysis on OHQ composite score by region in Study 302

	Droxidopa			Placebo		
Region	N	Mean	STD	N	Mean	STD
US	25	0.48	2.32	32	1.56	2.59
Non US	25	-0.12	1.95	19	0.46	1.88

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The sponsor conducted two phase III trials. Study 302 failed to meet its primary endpoint with p-value of 0.68 indicating no signal. The sponsor subsequently did a post-hoc analysis using the OHQ composite score as an additional assessment of the effect of droxidopa. The significant result of the OHQ composite score appeared to be driven by the OHDAS subscale. The seemingly significant result based on this post-hoc analysis led to the primary endpoint change in Study 301.

In Study 301, the primary efficacy variable was changed from the original endpoint of Item 1 of the OHSA to the OHQ composite score. The change was reflected on the protocol dated December 2009. Despite the primary endpoint change, the results in Study 301 appeared to be consistent. Analyses on OHQ composite score (primary endpoint), OHSA item 1 score (original primary endpoint), 3 minutes post-standing SBP and a number of sensitivity analyses all showed statistical significance favoring droxidopa group.

Even thought OHQ composite score appeared to be significant in both studies, it remains a concern whether OHQ composite score is a valid measurement for the NOH symptom. It is also unknown what effect size we should observe in OHQ composite score to show real clinical benefit. In addition, the analysis on OHQ score was post-hoc and the significant result of the OHQ composite score was driven by the OHDAS subscale. These led to the conclusion that only one of the two phase III trials demonstrated the efficacy of droxidopa.

5.2 Conclusions and Recommendations

The Chelsea-sponsored droxidopa clinical development program includes two randomized, placebo-controlled, double-blind studies (Studies 301 and 302). Only one study showed the efficacy of droxidopa.

Additional study is needed to confirm the finding.

CHECK LIST

Number of Pivotal Studies: 2

Trial Specification

Protocol Number (s): Droxidopa-301
Protocol Title (optional):
Phase: 3
Control: Placebo Control
Blinding: Double-Blind
Number of Centers: 65
Region(s) (Country): US, Ukraine, Canada, Germany, France, Italy, Austria, Cztech, Romania
Duration: 7 days
Treatment Arms: Placebo
Treatment Schedule: treated with individually optimized dose from titration stage
Randomization: Yes
Ratio: 1:1
Primary Endpoint: OHQ composite score
Primary Analysis Population: mITT
Statistical Design: Superiority
Primary Statistical Methodology: ANCOVA
Interim Analysis: No
Sample Size: 162
Sample Size Determination: Was it calculated based on the primary endpoint variable and the analysis being used for the primary variable?

Statistic = chi-square
Power= 80%
Δ= 1.2/2.39 = 0.5
α = 0.05

- Was there an **Alternative Analysis** in case of violation of assumption? No
- Were there any major changes, such as changing the statistical analysis methodology or changing the primary endpoint variable? Yes
- Were the **Covariates** pre-specified in the protocol? Yes
- Did the Applicant perform **Sensitivity Analyses**? Yes
- How were the **Missing Data** handled? LOCF, excluding missing data
- Was there a **Multiplicity** involved? Yes

  If yes,

  Multiple Arms (Yes/No)? No
  Multiple Endpoints (Yes/No)? Yes
  Which method was used to control for type I error? Sequential testing

- **Multiple Secondary Endpoints**: Yes, in the label. Sequential testing was used.

Were Subgroup Analyses Performed (Yes/No)? Yes

- Were there any **Discrepancies** between the protocol/statistical analysis plan vs. the study report? No
- Overall, was the study positive (Yes/No)? Yes

Trial Specification

Protocol Number (s): Droxidopa-302
Protocol Title (optional):
Phase: 3
Control: Placebo Control
Blinding: Double-Blind
Number of Centers: 71
Region(s) (Country): US, UK, Poland, New Zealand, Australia, Canada
Duration: 14 days
Treatment Arms: Placebo
Treatment Schedule: treated with individually optimized dose from titration stage
Randomization: Yes
Ratio: 1:1
Primary Endpoint: OHSA item 1 score
Primary Analysis Population: ITT
Statistical Design: Superiority
Primary Statistical Methodology: Wilcoxon Rank Sum Test
Interim Analysis: No
Sample Size: 101
Sample Size Determination: Was it calculated based on the primary endpoint variable and the analysis being used for the primary variable?

Statistic = z statistic
Power= 80%
Δ= 1.6/2.5 = 0.64
α = 0.05

- Was there an **Alternative Analysis** in case of violation of assumption? No
- Were there any major changes, such as changing the statistical analysis methodology or changing the primary endpoint variable? No
- Were the **Covariates** pre-specified in the protocol? N/A
- Did the Applicant perform **Sensitivity Analyses**? Yes
- How were the **Missing Data** handled? LOCF, excluding missing data
- Was there a **Multiplicity** involved? Yes

If yes,
Multiple Arms (Yes/No)? No
Multiple Endpoints (Yes/No)? Yes

Which method was used to control for type I error? Sequential testing for multiple secondary endpoints

- **Multiple Secondary Endpoints**: Not significant

Were Subgroup Analyses Performed (Yes/No)? Yes

- Were there any **Discrepancies** between the protocol/statistical analysis plan vs. the study report? No
- Overall, was the study positive (Yes/No)? No

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

/s/

JIALU ZHANG
01/23/2012

HSIEN MING J HUNG
01/23/2012

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 203-202 **Applicant: Chelsea** **Stamp Date: 09/28/2011**

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

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

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

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? __Yes_____

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

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

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	x			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	x			Validity of the primary endpoint may be a review issue
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			x	The interim analysis was only for safety data.
Appropriate references for novel statistical methodology (if present) are included.			x	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.			x	

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	x			

Reviewing Statistician Date

Supervisor/Team Leader Date

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

/s/

JIALU ZHANG
11/09/2011

HSIEN MING J HUNG
11/10/2011