

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

209241Orig1s000

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
STATISTICAL REVIEW AND EVALUATION
CLINICAL STUDIES

NDA #: 209,241

Drug Name: INGREZZA (valbenazine tosylate) 40 mg capsules

Indications: Treatment of Tardive Dyskinesia

Applicant: Neurocrine Biosciences

Dates: Submitted: 08/11/2016
PDUFA due date: 04/11/2017

Review Priority: Priority

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Thomas Birkner, Ph.D.

Concurring Reviewers: Peiling Yang, Ph.D.
H.M. James Hung, Ph.D.

Medical Division: Division of Psychiatry Products

Clinical Team: Michael Davis, MD; Javier Muniz, MD (TL); Brian Miller, MD

Project Manager: Jasmeet (Mona) Kalsi

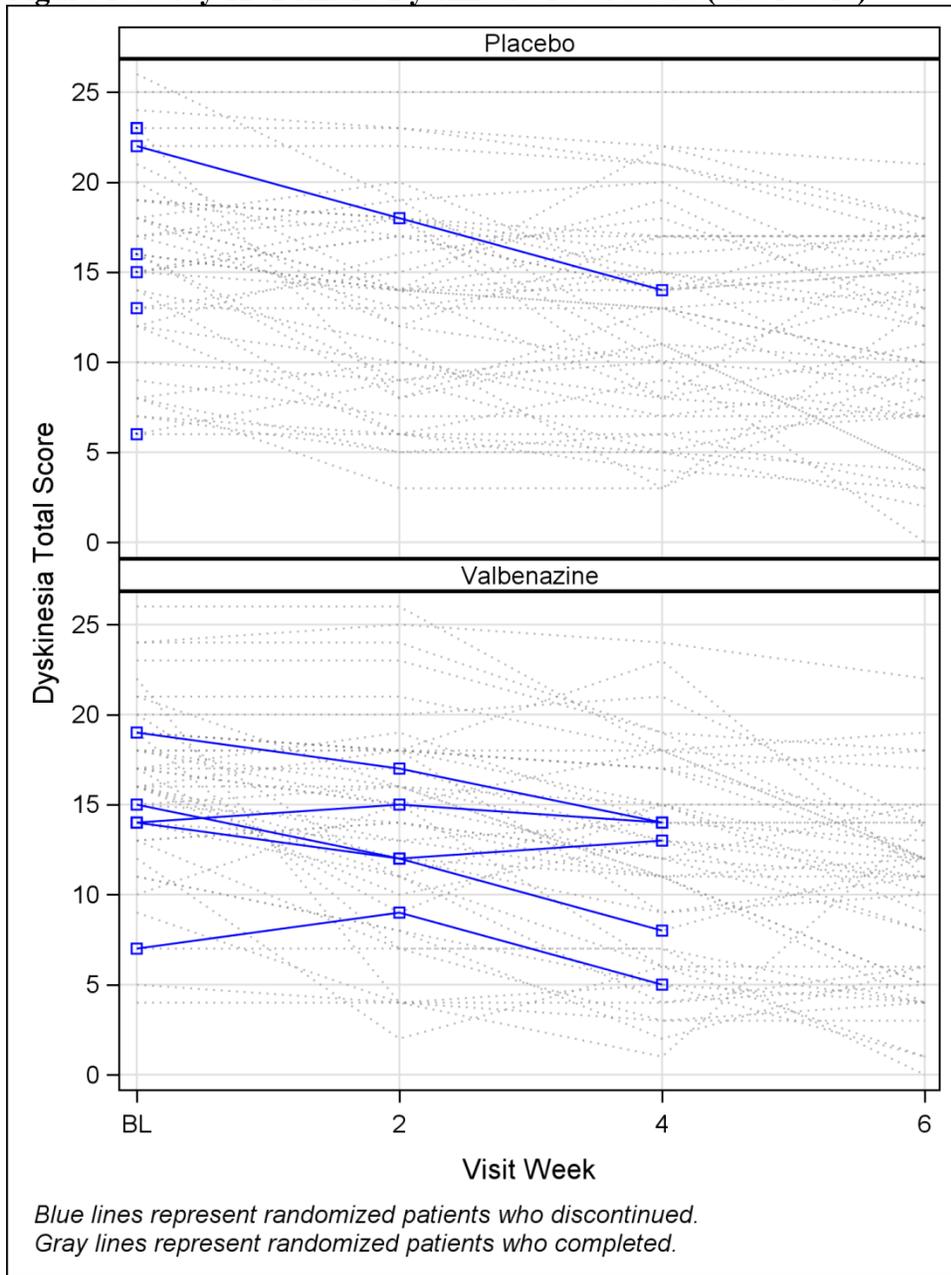
Keywords: mixed models, multiple endpoints, tipping point

This addendum to the original statistical review of NDA 209,241 has been prepared to document the individual patient response on the primary efficacy measure (AIMS dyskinesia total score) over time for both studies 1202 and 1304. This addendum does not change the conclusions of the original statistical review filed on 03/06/2017.

The figures include all randomized subjects with at least an AIMS dyskinesia total score value at baseline. Scores for patients who discontinued for any reason during the double-blind period are displayed as blue squares (connected by a blue line to create the profile over time). Gray dotted lines represent the AIMS dyskinesia total score trajectories for the completers.

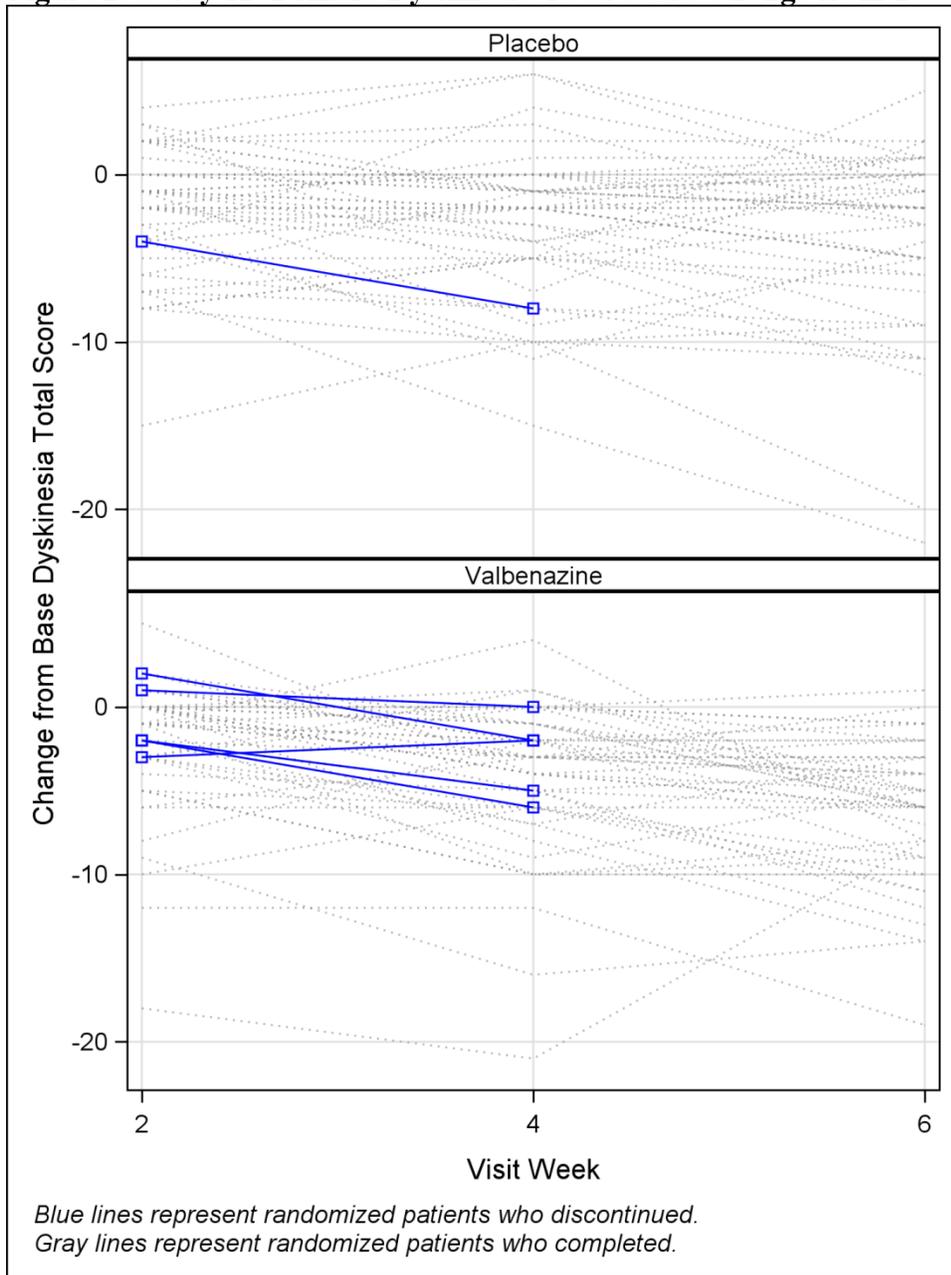
Note that in Study 1202 central ratings of the video recorded AIMS assessments were only performed at baseline and week 6. In order to display informative patient efficacy profiles for Study 1202 the AIMS scores generated by the independent site raters are used, which are also available for weeks 2 and 4 of the double-blind period.

Figure 1. Study 1202 AIMS Dyskinesia Total Score (Site raters)



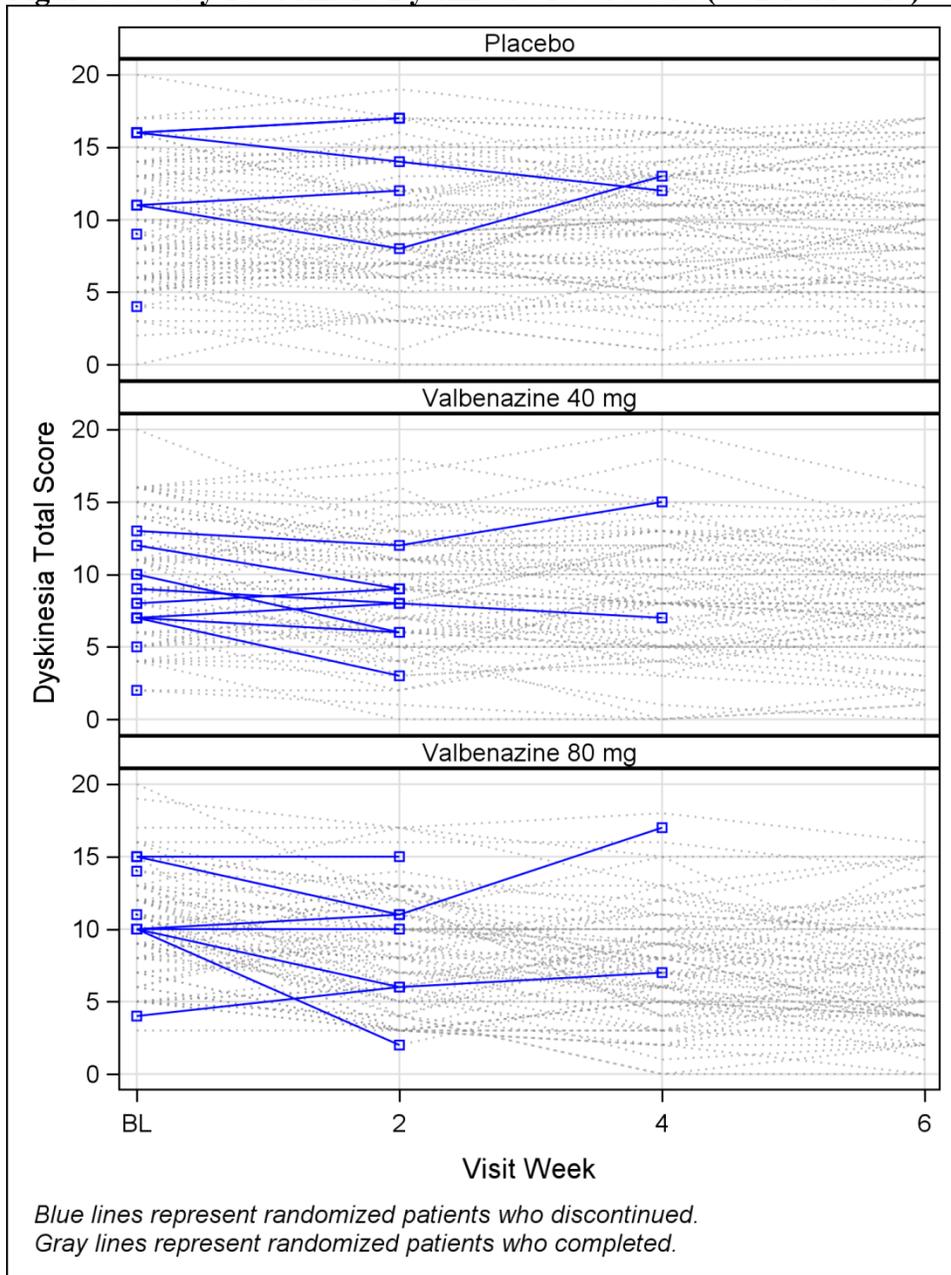
(Source: Reviewer; BL=Baseline)

Figure 2. Study 1202 AIMS Dyskinesia Total Score Change from Baseline (Site raters)



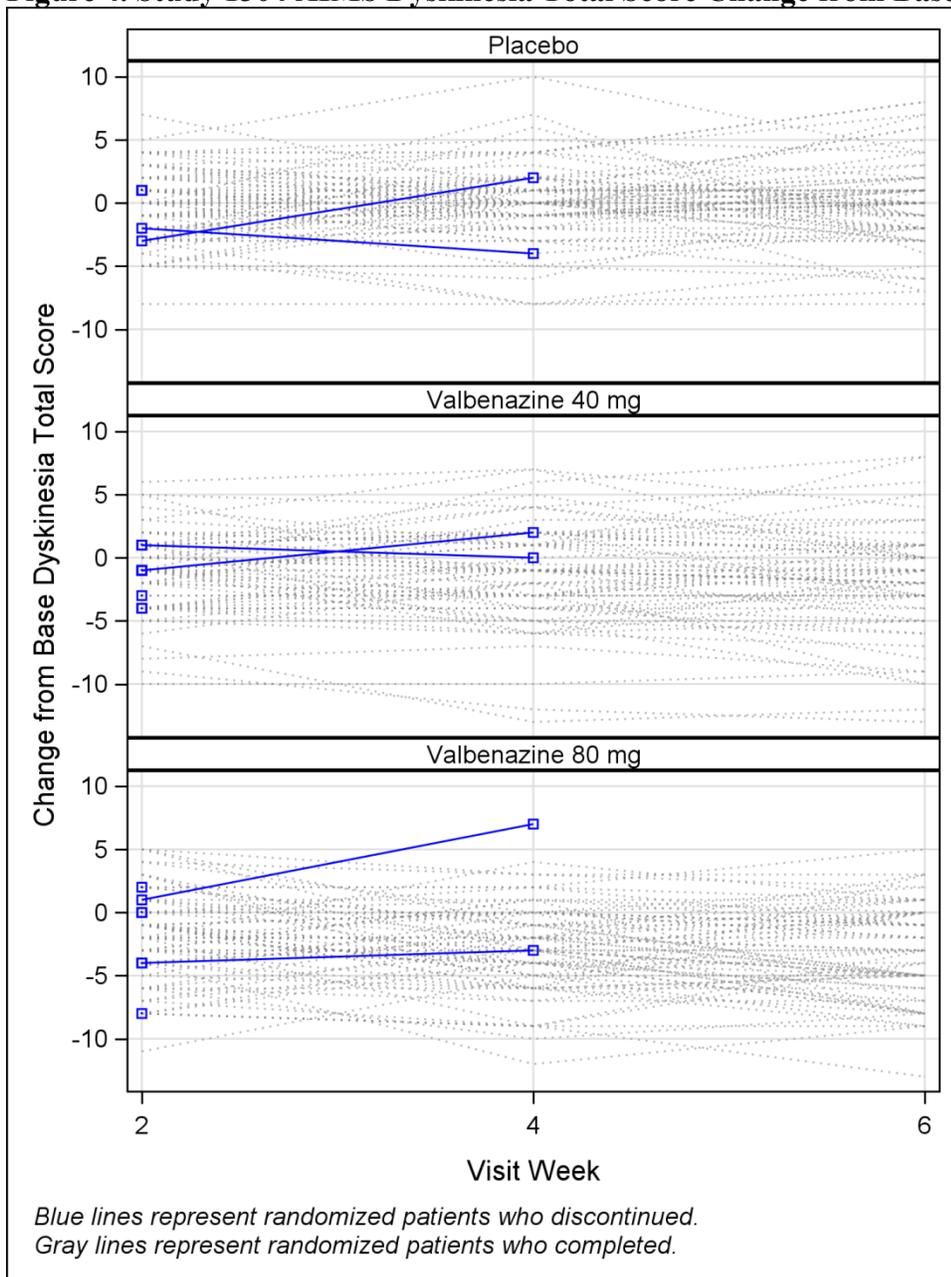
(Source: Reviewer)

Figure 3. Study 1304 AIMS Dyskinesia Total Score (Central raters)



(Source: Reviewer; BL = Baseline)

Figure 4. Study 1304 AIMS Dyskinesia Total Score Change from Baseline (Central raters)



(Source: Reviewer)

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

/s/

THOMAS BIRKNER
03/16/2017

PEILING YANG
03/16/2017

HSIEN MING J HUNG
03/16/2017

C – Carcinogenicity Studies Statistical Review and Assessment

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

Statistical Review and Evaluation
CARCINOGENICITY STUDIES

IND/NDA Number: NDA-209241
Drug Name: NBI-98854 (b) (4)
Indication: Treatment of Tardive Dyskinesia
Studies: 104 Week Carcinogenicity Studies in Rats and 26 Weeks in Mice
Applicant: Sponsor:
Neurocrine Biosciences
12780 El Camino Real
San Diego, California 92130
United States of America
Testing Facility for Rats: (b) (4)
Testing Facility for Mice: (b) (4)
Documents Reviewed: Electronic submission: Submitted on December 15, 2016
Electronic data: Submitted on December 15, 2016
Review Priority: Standard
Biometrics Division: Division of Biometrics - VI
Statistical Reviewer: Hepei Chen
Concurring Reviewer: Karl Lin, Ph.D.
Medical Division: Division of Psychiatry Products
Reviewing Pharmacologist: Darren Fegley, Ph.D.
Keywords: Carcinogenicity, Dose response

Table of Contents

1.	Background.....	3
2.	Rat Study.....	3
2.1.	Sponsor's analyses.....	4
2.1.1.	Survival analysis.....	4
	Sponsor's findings.....	4
2.1.2.	Tumor data analysis.....	4
	Adjustment for multiple testing.....	5
	Sponsor's findings.....	5
2.2.	Reviewer's analyses.....	5
2.2.1.	Survival analysis.....	5
	Reviewer's findings.....	5
2.2.2.	Tumor data analysis.....	6
	Multiple testing adjustment.....	6
	Reviewer's findings.....	7
3.	Mouse Study.....	8
3.1.	Sponsor's analyses.....	8
3.1.1.	Survival analysis.....	8
	Sponsor's findings.....	9
3.1.2.	Tumor data analysis.....	9
	Sponsor's findings.....	9
3.2.	Reviewer's analyses.....	9
3.2.1.	Survival analysis.....	10
	Reviewer's findings.....	10
3.2.2.	Tumor data analysis.....	10
	Reviewer's findings.....	10
4.	Summary.....	10
5.	Appendix.....	13
	Table 1A: Intercurrent mortality rate in male rats	
	Table 1B: Intercurrent mortality rate in female rats	
	Table 2A: Tumor rates and p-values for trend and pairwise comparisons in male rats	
	Table 2B: Tumor rates and p-values for trend and pairwise comparisons in female rats	
	Table 3A: Intercurrent mortality rate in male mice	
	Table 3B: Intercurrent mortality rate in female mice	
	Table 4A: Tumor rates and p-values for trend and pairwise comparisons in male mice	
	Table 4B: Tumor rates and p-values for trend and pairwise comparisons in female mice	
	Figure 1A: Kaplan-Meier survival functions for male rats	
	Figure 1B: Kaplan-Meier survival functions for female rats	
	Figure 2A: Kaplan-Meier survival functions for male mice	
	Figure 2B: Kaplan-Meier survival functions for female mice	
6.	References.....	30

1. Background

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were to evaluate the carcinogenic potential and determine the toxicokinetics of the test article, NBI-98854, when administered daily via oral gavage to rats for the intended duration of 104 weeks, and to mice for 26 weeks.

In this review the phrase "dose response relationship" refers to the linear component (trend) of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

2. Rat Study

Two separate experiments, one in male rats and one in female rats were conducted. As indicated in Table 1, in each of these two experiments there were three treated groups and one vehicle control group. Two hundred forty Crl:CD(SD) rats of each sex were assigned randomly to the treated and control groups in equal size of 60 rats per group. The dose levels for treated groups were 0.5, 1, and 2 mg/kg/day for both male and female rats. In this review these dose groups were referred to as the low (Group 2), mid (Group 3), and high (Group 4) dose groups, respectively. The rats in the vehicle control group were administered with the vehicle [0.25% (w/v) methylcellulose (4000 cPs) in reverse osmosis water], and handled for the same duration and in the same manner as the treated groups.

Table 1: Experimental Design in Rat Study

Group No.	No. of Toxicity Animals		Test Material	Dosage Level (mg/kg/day)	
	Male	Female		Male	Female
1	60	60	Vehicle control	0	0
2	60	60	NBI-98854 low	0.5	0.5
3	60	60	NBI-98854 mid	1	1
4	60	60	NBI-98854 high	2	2

Toxicokinetic and carcinogenicity animals were checked twice daily (a.m. and p.m.) for mortality, abnormalities, and signs of pain or distress. Detailed observations were conducted for carcinogenicity animals once during the predose phase, prior to dosing on Day 1, and weekly (based on Day 1) throughout the dosing phase. Detailed observations were also collected on days of scheduled sacrifice (animals scheduled for sacrifice only). Scheduled study termination was planned for Week 104 of the dosing phase. Three early scheduled terminations occurred based on survival.

- Due to Group 1 males having reached 20 surviving animals on Day 633 of the dosing phase, all surviving males were sacrificed.
- Due to Group 2 females having reached 15 surviving animals on Day 636 of the dosing phase, all surviving females in Group 2 were sacrificed.
- Due to Group 1 females having reached 20 surviving animals on Day 641 of the dosing phase, all surviving females in Groups 1, 3, and 4 were sacrificed.

2.1. Sponsor's analyses

2.1.1. Survival analysis

The sponsor performed the tests to compare survival with a two-sided risk for increasing and decreasing mortality with dose. Tests were performed for dose response and for each treated group against vehicle control using Kaplan-Meier product-limit estimation curves, along with log-rank and Wilcoxon tests, using the LIFETEST procedure in SAS. The time to death or sacrifice (in weeks) was the dependent variable. Treatment group was included as the strata. Animals with a death or sacrifice status recorded as a planned sacrifice (interim or terminal) or an accidental death were censored in the analysis.

Sponsor's findings:

The sponsor's analysis showed that the numbers of rats surviving to their terminal necropsy were 20 (33.33%), 22 (36.67%), 23 (38.33%), and 33 (55.00%) in Groups 1, 2, 3, and 4 for male rats, respectively, and 20 (33.33%), 15 (25.00%), 18 (30.00%), and 27 (45.00%) for female rats respectively. The sponsor reported that for males, the high dose group (2 mg/kg/day) had lower mortality than the vehicle control group (27/60 versus 40/60 in the vehicle control group), with $p=0.0216$ and $p=0.0395$ for the Log-Rank and Wilcoxon tests respectively. The dose response was also significant, with $p=0.0195$ and $p=0.0394$ for the Log-Rank and Wilcoxon tests respectively. For females, no significant findings were noted in the sponsor's report.

2.1.2. Tumor data analysis

The sponsor analyzed those tumors from tissues that were listed in the protocol to be examined. For each given tumor type, statistical analysis was performed if the incidence at least one treated group was increased by at least two occurrences over the vehicle control group. Tests to compare tumor incidence were performed with a one-sided risk for increasing incidence with dose. Tests were performed for dose response and for each treated group against vehicle control.

For tumors occurring in animals dying spontaneously or sacrificed in extremis during the study, the pathologist classified the context of observation as one of the following:

- (1) Fatal: the tumor was a factor in the demise of the animal.
- (2) Non-fatal: the tumor was not a factor in the demise of the animal.
- (3) Uncertain

Occult or non-palpable tumors were analyzed by the IARC asymptotic fixed interval based prevalence test (Peto et al., 1980). The cut off points for the interval based test were Weeks 0 to 52, 53 to 78, 79 to before terminal sacrifice, and the terminal sacrifice. Fatal and non-fatal tumors were analyzed together, with separate strata for each. There were no tumors of uncertain context. The test was implemented using PROC MULTTEST in the SAS system. In the case of sparse tables (<10 total in the strata), the exact form of the test was used for that strata. Otherwise, the asymptotic version of the test was used. Animals were assigned to the terminal sacrifice strata based on the death or sacrifice status recorded in the data, and were not assigned based on their week of necropsy.

Observable or palpable (superficial as in mammary or skin) tumors were analyzed using the methods previously described for analyzing survival, using the time to death or time of detection of the tumor (in weeks) as a surrogate for the tumor onset time. Comparisons between vehicle and treated groups were performed with a one-sided risk for increasing incidence with dose.

Unadjusted P-values were reported for tumors. Where applicable, site or tumor combinations were statistically analyzed if the incidence in at least one treated group was increased by at least two occurrences compared to the vehicle control group. The criteria for combination were based on the work of McConnell et al. (McConnell et al., 1986) and as indicated by the study pathologist. Incidences of multiple-organ and combined neoplastic findings, such as hemangioma, fibrosarcoma, and endometrial stromal polyp were counted by animal, not by tissue type. Due to individual values being rounded for inclusion in the report, calculation of summary statistics from these reported values may, in some cases, yield minor differences.

Adjustment for multiple testing:

Indication of a possible treatment effect was assessed on the basis of rare or common tumor type, in line with the current FDA guidelines (FDA Draft Guidance for Industry, 2001).

Sponsor's findings:

For male rats, the sponsor reported statistically significant increases for the incidence of malignant fibrosarcoma in skin/subcutis at the low dose group when compared to the vehicle control group (p-value=0.0143 and 0.0156 for Log-rank test and Wilcoxon test, respectively), for the incidence of benign trichoepithelioma in skin/subcutis at the mid dose group when compared to the vehicle control group (p-value=0.0227 and 0.0265 for Log-rank test and Wilcoxon test, respectively), and for the incidence of combined tumor including benign adenoma sebaceous, basal cell tumor, keratoacanthoma, papilloma squamous cell, trichoepithelioma, and malignant carcinoma sebaceous and carcinoma squamous cell in skin/subcutis at the mid dose group when compared to the vehicle control group (p-value=0.0290 and 0.0457 for Log-rank test and Wilcoxon test, respectively).

No other statistically significant tumor findings were noted in both male and female rats.

2.2. Reviewer's analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing toxicologist, this reviewer independently performed the survival and tumor data analyses using the data provided by the sponsor electronically.

2.2.1. Survival analysis

The survival distributions of rats in all four groups (Groups 1, 2, 3, and 4) were estimated using the Kaplan-Meier product limit method. The dose response relationship was tested across Groups 1, 2, 3, and 4 using the likelihood ratio test, and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1A and 1B in the appendix for all four groups in male and female rats, respectively. The intercurrent mortality data of all four groups, and the results of the tests for dose response relationship and homogeneity of survivals for Groups 1, 2, 3, and 4 are given in Tables 1A and 1B in the appendix for male and female rats, respectively.

Reviewer's findings:

The reviewer's analysis showed that the numbers of rats surviving to their terminal necropsy were 20 (33.33%), 22 (36.67%), 23 (38.33%), and 33 (55.00%) in Groups 1, 2, 3, and 4 for male

rats, respectively, and 20 (33.33%), 15 (25.00%), 18 (30.00%), and 27 (45.00%) for female rats respectively. A statistically significant positive dose-response relationship in mortality was noted in male rats (p -value=0.0190), along with a statistically significant increase in the high dose group when compared to the vehicle control group (p -value=0.0227 and 0.0216 for the dose-response test and the log-rank test, respectively). No statistically significant findings in mortality were noted in female rats.

2.2.2. Tumor data analysis

The tumor data were analyzed for dose response relationships across Groups 1, 2, 3, and 4, and pairwise comparisons of each of the three treated groups (Groups 2, 3, and 4) against the vehicle control group (Group 1), using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993).

In the poly-k method, the adjustment for differences in mortality among treatment groups is made by modifying the number of animals at risk in the denominators in the calculations of overall tumor rates in the Cochran-Armitage test to reflect less-than-whole-animal contributions for animals that die without tumor before the end of the study (Bailer and Portier 1988). The modification is made by defining a new number of animals at risk for each treatment group. The number of animals at risk for the i -th treatment group R^*_i is defined as $R^*_i = \sum w_{ij}$ where w_{ij} is the weight for the j -th animal in the i -th treatment group, and the sum is over all animals in the group.

Bailer and Portier (1988) proposed the weight w_{ij} as follows:

$w_{ij} = 1$ to animals dying with the tumor, and
 $w_{ij} = (t_{ij} / t_{sacr})^k$ to animals dying without the tumor,

where t_{ij} is the time of death of the j -th animal in the i -th treatment group, and t_{sacr} is the planned (or intended) time of terminal sacrifice. The above formulas imply that animals living up to the end of the planned terminal sacrifice date without developing any tumor will also be assigned $w_{ij} = 1$ since $t_{ij} = t_{sacr}$.

Certain treatment groups of a study or the entire study may be terminated earlier than the planned (or intended) time of terminal sacrifice due to excessive mortalities. However, based on the principle of the Intention-to-treat (ITT) analysis in randomized trials, the t_{sacr} should not be affected by the unplanned early terminations. The t_{sacr} should always be equal to the planned (or intended) time of terminal sacrifice. For those animals that were sacrificed later than t_{sacr} , regardless their actual terminal sacrifice time, t_{sacr} was used as their time of terminal sacrifice in the analysis.

One critical point for Poly-k test is the choice of the appropriate value of k , which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of $k=3$ is suggested in the literature. Hence, this reviewer used $k=3$ for the analysis of this data.

Multiple testing adjustment:

For the adjustment of multiple testing, this reviewer used the methodologies suggested in the FDA guidance for statistical design and analysis of carcinogenicity studies (2001). For dose response relationship tests, the guidance suggests the use of test levels of $\alpha=0.01$ for common

tumors and $\alpha=0.05$ for rare tumors for a submission with one two-year study in one species and one short-term study with another species, in order to keep the overall false-positive rate at the nominal level of approximately 10%. For multiple pairwise comparisons of treated group with control group, however, the guidance indicated that the corresponding multiple testing adjustment is still under development and not yet available. To be conservative, the test level of $\alpha=0.05$ was used for pairwise comparisons of treated group with control group for both rare and common tumors in this study.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman (1998). In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Rahman and Lin (2008) showed that this rule for multiple testing for dose response relationship is also suitable for Poly-k tests.

A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. However, if the background information for the common or rare tumor is not available, the number of animals bearing tumors in the vehicle control group in the present study was used to determine the common or rare tumor status in the review report. That is, if the number of animals bearing tumors in the vehicle control group is 0, then this tumor is considered as the rare tumor; otherwise, if the number of animals bearing tumors in the control group is greater than or equal to 1, then this tumor is considered as the common tumor.

Reviewer's findings:

The tumor rates and the p-values of the tested tumor types are listed in Tables 2A and 2B in the appendix for male and female rats, respectively. The tumor types with p-values less than or equal to 0.05 for dose response relationship and/or pairwise comparisons of treated groups and vehicle control are reported in Table 2.

Table 2. Summary Table of Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship and/or Pairwise Comparisons of Treated Groups and Vehicle Control Group in Rats

Organ name	Tumor name	0 mg	5 mg	10 mg	20 mg
		Vehicle (C) P - Trend	Low (L) P - L vs. C	Mid (M) P - M vs. C	High (H) P - H vs. C
Male-Skin/Subcutis	M-Fibrosarcoma	0/58 (26) 0.5269	5/59 (31) 0.0406 \$	0/59 (27) NC	2/59 (32) 0.3001
Female Mammary Gland	B-Adenoma	1/59 (28) 0.0394 @	0/59 (27) 0.4909	0/60 (25) 0.4717	4/60 (34) 0.2437

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

\$ = Statistically significant at 0.05 level in rare tumor for test of pairwise comparisons;

@ = Not statistically significant at 0.01 level in common tumor for test of dose response relationship;

NC = Not calculable.

Based on the criteria of adjustment for multiple testing discussed above, the reviewer's analysis showed that a statistically significant increase ($p = 0.0406$) for the incidence rates of malignant fibrosarcoma of skin/subcutis in the low dose group when comparing to the vehicle control group in male rats, if this tumor was considered to be rare. A p-value of 0.0394 was noted for the dose response relationship of the benign adenoma in mammary gland for female rats. However, this trend was not statistically significant as this tumor was considered to be common. No other

statistically significant findings were noted for male and female rats.

3. Mouse Study

Two separate experiments, one in male mice and one in female mice were conducted. As indicated in Table 3, in each of these two experiments there were three treated groups, one vehicle control group, and one positive control group. One hundred and ten hemizygous Tg.rasH2 mice of each sex were assigned randomly in size of 25 mice per group to the treated and vehicle control groups, and 10 mice to the positive control group. The dose levels for treated groups were 10, 30, and 75 mg/kg/day for both male and female mice. In this review these dose groups were referred to as the low (Group 2), mid (Group 3), and high (Group 4) dose groups, respectively. The mice in the vehicle control and the positive control group were administered with the vehicle control [0.25% methylcellulose (4000 cps) in de-ionized (DI) water] and the positive control [Urethane in 0.9% NaCl (saline)], respectively, and handled for the same duration and in the same manner as the treated groups.

Table 3. Experimental Design in Mouse Study

Group No.	No. of Toxicity Animals		Test Material	Dosage Level (mg/kg/day)	
	Male	Female		Male	Female
1	25	25	Vehicle control	0	0
2	25	25	NBI-98854 low	10	10
3	25	25	NBI-98854 mid	30	30
4	25	25	NBI-98854 high	75	75
5	10	10	Positive control	0	0

All animals were observed twice daily at least 6 hours apart for moribundity and mortality, except as noted (see Deviations). For the Main Cohort only, cage side observations were performed daily within 2 hours after the last animal was dosed in each group. For the Main Cohort animals, detailed hands-on observations were performed on Day 1 and weekly thereafter (at the time animals were weighed). Any Main Cohort animals found dead were necropsied as soon as possible after discovery, usually within 8 hours. Moribund animals were sacrificed by CO₂ overdose and carcasses were refrigerated until necropsied, when needed. In the Main Cohort (Groups 1-4), surviving animals were sacrificed by CO₂ overdose on Day 183 or Day 184 and necropsied. Prior to sacrifice, animals were weighed to the nearest 0.1 gram. All surviving animals in the positive control (Group 5) were sacrificed by CO₂ overdose on Day 64. A complete necropsy was performed for all Main Cohort animals (Groups 1-4); in addition, animals found dead or moribund sacrificed were also evaluated for evidence of gavage error.

3.1. Sponsor's analyses

3.1.1. Survival analysis

The sponsor calculated the Kaplan-Meier estimates of group survival rates by sex and showed in graph. The generalized Wilcoxon test for survival was used to compare the homogeneity of survival rates across the vehicle control and test article groups, by sex, at the 0.05 significance level. If the survival rates were significantly different, the generalized Wilcoxon test was used to make pairwise comparisons of each test article group with the vehicle control group. Additionally, the positive control group was compared to the vehicle control group using the

generalized Wilcoxon test. Survival times in which the status of the animal's death was classified as an accidental death, planned interim sacrifice or terminal sacrifice were considered censored values for the purpose of the Kaplan-Meier estimates and survival rate analyses.

Sponsor's findings:

The sponsor's analysis showed that the numbers of mice surviving to their terminal necropsy were 25 (100%), 25 (100%), 23 (92%), and 24 (96%) in Groups 1, 2, 3, and 4 for male, respectively, and 25 (100%), 25 (100%), 25 (100%), and 23 (92%) for female, respectively. The sponsor reported no statistically significant findings in survival rates for male and female mice.

3.1.2. Tumor data analysis

For the vehicle and treated groups, the sponsor used Peto's mortality-prevalence method to analyze the incidence of tumors, incorporating the context (incidental, fatal, or mortality independent) in which tumors were observed without continuity correction. The following fixed intervals were used for incidental tumor analyses: Days 1 through 130, and Days 131 through and including terminal sacrifice. A minimum exposure of 130 days was considered sufficient to be included with animals surviving through scheduled termination. All tumors in the scheduled terminal sacrifice interval were considered incidental for the purpose of statistical analysis. Tumors classified as mortality-independent were analyzed with Peto's mortality independent method incorporating the day of detection. A 1-sided comparison of each test article group with the vehicle control was performed. An exact permutation test was conducted for all analyses. Findings were evaluated for statistical significance at both the 0.01 and 0.05 levels and all p values were reported.

Each diagnosed tumor type was analyzed separately and, at the discretion of the study director, analysis of combined tumor types and/or organs was performed. All metastases and invasive tumors were considered secondary and not statistically analyzed.

For the vehicle and positive control groups, because the positive control group was scheduled for early terminal sacrifice, tumor incidence in the positive control group was compared to the vehicle control group with a 1-sided Fisher's exact test at both the 0.01 and 0.05 significance levels and all p values were reported. Only the following tumors were statistically analyzed: alveolarbronchiolar adenoma, alveolar-bronchiolar carcinoma, and hemangiosarcoma in the spleen.

Sponsor's findings:

The sponsor's analysis showed that for both male and female mice, there were no statistically significant tumor findings in the treated groups when compared to the vehicle control group; while statistically significant increases in the incidence of alveolar-bronchiolar adenoma in lungs with bronchi and hemangiosarcoma in spleen were noted when comparing the positive control with the vehicle control group for both male and female rats.

3.2. Reviewer's analyses

Similar to the rat study, this reviewer independently performed survival and tumor data analyses of mouse data to verify sponsor's analyses. Data used in this reviewer's analyses were provided by the

sponsor electronically.

For the analysis of both the survival data and the tumor data, this reviewer used similar methodologies that were used for the analyses of the rat survival and tumor data.

3.2.1. Survival analysis

The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 2A and 2B in the appendix for male and female mice, respectively. The intercurrent mortality data, and the results of the tests for dose response relationship and homogeneity of survivals for the combined vehicle control, low, mid, and high dose groups were given in Tables 3A and 3B in the appendix for male and female mice, respectively.

Reviewer's findings:

In the reviewer's analysis, the numbers of mice surviving to their terminal necropsy were 25 (100%), 25 (100%), 23 (92%), and 24 (96%) in Groups 1, 2, 3, and 4 for male, respectively, and 25 (100%), 25 (100%), 25 (100%), and 23 (92%) for female, respectively. A statistically significant positive dose-response relationship in mortality was noted in female mice (p -value=0.0179), without any statistically significant pairwise comparisons between the vehicle control groups and the treated groups. No statistically significant findings in mortality were noted in male mice.

3.2.2. Tumor data analysis

The tumor rates and the p -values of the tested tumor types are given in Tables 4A and Table 4B in the appendix, for male and female mice, respectively.

Reviewer's findings:

The reviewer's analysis showed no statistically significant dose response relationship or pairwise comparisons in the treated groups when compared to the vehicle control group for both male and female mice. When comparing the positive control with the vehicle control group, statistically significant increases in the incidence of alveolar-bronchiolar adenoma in lungs with bronchi and hemangiosarcoma in spleen were noted for both male and female rats (all p -values < 0.0001).

4. Summary

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were to evaluate the carcinogenic potential and determine the toxicokinetics of the test article, NBI-98854, when administered daily via oral gavage to rats for the intended duration of 104 weeks, and to mice for 26 weeks.

Rat Study:

Two separate experiments, one in male rats and one in female rats were conducted. In each of these two experiments there were three treated groups and one vehicle control group. Two hundred forty CrI:CD(SD) rats of each sex were assigned randomly to the treated and control

groups in equal size of 60 rats per group. The dose levels for treated groups were 0.5, 1, and 2 mg/kg/day for both male and female rats.

The reviewer's analysis showed that the numbers of rats surviving to their terminal necropsy were 20 (33.33%), 22 (36.67%), 23 (38.33%), and 33 (55.00%) in Groups 1, 2, 3, and 4 for male rats, respectively, and 20 (33.33%), 15 (25.00%), 18 (30.00%), and 27 (45.00%) for female rats respectively. A statistically significant positive dose-response relationship in mortality was noted in male rats (p -value=0.0190), along with a statistically significant increase in the high dose group when compared to the vehicle control group (p -value=0.0227 and 0.0216 for the dose-response test and the log-rank test, respectively). No statistically significant findings in mortality were noted in female rats.

Based on the criteria of adjustment for multiple testing discussed above, the reviewer's analysis showed that a statistically significant increase ($p = 0.0406$) for the incidence rates of malignant fibrosarcoma of skin/subcutis in the low dose group when comparing to the vehicle control group in male rats, if this tumor was considered to be rare. A p -value of 0.0394 was noted for the dose response relationship of the benign adenoma in mammary gland for female rats. However, this trend was not statistically significant as this tumor was considered to be common. No other statistically significant findings were noted for male and female rats.

Mouse Study:

Two separate experiments, one in male mice and one in female mice were conducted. In each of these two experiments there were three treated groups, one vehicle control group, and one positive control group. One hundred and ten hemizygous Tg.rasH2 mice of each sex were assigned randomly in size of 25 mice per group to the treated and vehicle control groups, and 10 mice to the positive control group. The dose levels for treated groups were 10, 30, and 75 mg/kg/day for both male and female mice.

In the reviewer's analysis, the numbers of mice surviving to their terminal necropsy were 25 (100%), 25 (100%), 23 (92%), and 24 (96%) in Groups 1, 2, 3, and 4 for male, respectively, and 25 (100%), 25 (100%), 25 (100%), and 23 (92%) for female, respectively. A statistically significant positive dose-response relationship in mortality was noted in female mice (p -value=0.0179), without any statistically significant pairwise comparisons between the vehicle control groups and the treated groups. No statistically significant findings in mortality were noted in male mice.

The reviewer's analysis showed no statistically significant dose response relationship or pairwise comparisons in the treated groups when compared to the vehicle control group for both male and female mice. When comparing the positive control with the vehicle control group, statistically significant increases in the incidence of alveolar-bronchiolar adenoma in lungs with bronchi and hemangiosarcoma in spleen were noted for both male and female rats (all p -values < 0.0001).

Hepei Chen.
Mathematical Statistician

Concur: Karl Lin, Ph.D.
Team Leader, DBVI

Cc: Archival NDA 209241

Dr. Darren Fegley
Dr. Lillian Patrician

5. Appendix

Table 1A: Intercurrent Mortality Rate in Male Rats

Week / Type of Death	Vehicle Control		Low		Mid		High	
	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %
0 - 52	4	6.67	6	10.00	5	8.33	6	10.00
53 - 78	22	43.33	18	40.00	19	40.00	11	28.33
79 - 91	14	66.67	14	63.33	13	61.67	10	45.00
Terminal sacrifice	20	33.33	22	36.67	23	38.33	33	55.00
Total	60		60		60		60	
Test	All Dose Groups		Vehicle Control vs. Low		Vehicle Control vs. Mid		Vehicle Control vs. High	
Dose-Response (Likelihood Ratio)	0.0190*		0.6731		0.6586		0.0227*	
Homogeneity (Log-Rank)	0.1132		0.6702		0.6556		0.0216*	

#All Cum. % Cumulative Percentage except for Terminal sacrifice;

* = Significant at 5% level;

Table 1B: Intercurrent Mortality Rate in Female Rats

Week / Type of Death	Vehicle Control		Low		Mid		High	
	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %
0 - 52	5	8.33	3	5.00	7	11.67	4	6.67
53 - 78	22	45.00	23	43.33	23	50.00	12	26.67
79 - 91	13	66.67	19	75.00	12	70.00	15	51.67
92 - 104							2	3.33
Terminal sacrifice	20	33.33	15	25.00	18	30.00	27	45.00
Total	60		60		60		60	
Test	All Dose Groups		Vehicle Control vs. Low		Vehicle Control vs. Mid		Vehicle Control vs. High	
Dose-Response (Likelihood Ratio)	0.0534		0.4849		0.3846		0.0958	
Homogeneity (Log-Rank)	0.0480*		0.4770		0.3776		0.0909	

#All Cum. % Cumulative Percentage except for Terminal sacrifice;

* = Significant at 5% level;

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Rats

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Adrenal, Cortex	B-Adenoma	1/60 (28) 0.7560	1/60 (28) NC	1/60 (28) NC	0/60 (31) 0.5254
	M-Carcinoma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
	B-Adenoma/M-Carcinoma	1/60 (28) 0.4537	1/60 (28) NC	1/60 (28) NC	1/60 (31) 0.2718
	B-Hemangioma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
Adrenal, Medulla	B-Pheochromocytoma	4/60 (29) 0.8509	6/60 (31) 0.4102	5/60 (29) 0.5000	2/60 (32) 0.7108
	M-Malignant Pheochromocytoma	0/60 (27) 0.6277	2/60 (29) 0.2636	1/60 (28) 0.5091	0/60 (31) NC
	B-Pheochromocytoma/ M-Malignant Pheochromocytoma	4/60 (29) 0.8875	8/60 (31) 0.2013	6/60 (30) 0.3878	2/60 (32) 0.7108
Brain	B-Granular Cell Tumor	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
	M-Glioma	0/60 (27) 0.5259	1/60 (29) 0.5179	1/60 (28) 0.5091	0/60 (31) NC
Epididymis	M-Malignant Mesothelioma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
Femur	M-Osteosarcoma	0/60 (27) 0.5175	0/60 (28) NC	1/60 (28) 0.5091	0/59 (31) NC
Heart	M-Endocardial Schwannoma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
Hemolympho- Reticular System	M-Histiocytic Sarcoma	1/60 (28) 0.4291	2/60 (29) 0.5134	1/60 (28) NC	2/60 (32) 0.5508
	M-Malignant Lymphoma	1/60 (28) 0.4865	1/60 (29) 0.2544	0/60 (28) 0.5000	1/60 (32) 0.2802

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

**Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Kidney	B-Adenoma, Tubule Cell, Amp*	1/60 (28) 0.6265	3/60 (30) 0.3325	1/60 (28) NC	1/60 (31) 0.2718
	B-Adenoma, Tubule Cell, Amp*/ M-Carcinoma, Tubule C	1/60 (28) 0.6265	3/60 (30) 0.3325	1/60 (28) NC	1/60 (31) 0.2718
	B-Lipoma	0/60 (27) 0.5175	1/60 (28) 0.5091	0/60 (28) NC	0/60 (31) NC
	M-Carcinoma, Tubule Cell	0/60 (27) 0.5130	1/60 (29) 0.5179	0/60 (28) NC	0/60 (31) NC
	M-Carcinoma, Tubule Cell, A*	0/60 (27) 0.3405	1/60 (29) 0.5179	0/60 (28) NC	1/60 (31) 0.5345
	M-Carcinoma, Tubule Cell/ M-Carcinoma, Tubule Cell, A*	0/60 (27) 0.4850	2/60 (29) 0.2636	0/60 (28) NC	1/60 (31) 0.5345
	M-Liposarcoma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
	Liver	M-Carcinoma, Hepatocellular	2/60 (28) 0.9423	0/60 (28) 0.7545	0/60 (28) 0.7545
Lung	M-Squamous Cell Carcinoma	0/60 (27) 0.2783	0/59 (28) NC	0/60 (28) NC	1/60 (32) 0.5424
Lymph Node, Mesenteric	B-Hemangioma	0/60 (27) 0.4850	2/60 (29) 0.2636	0/60 (28) NC	1/60 (31) 0.5345
Mammary Gland	B-Fibroadenoma	1/57 (26) 0.6487	0/59 (27) 0.5094	1/59 (28) 0.2642	0/57 (29) 0.5273
	M-Carcinoma	3/57 (27) 0.9169	0/59 (27) 0.8821	2/59 (29) 0.5347	0/57 (29) 0.8945
Nerve, Sciatic	B-Schwannoma	0/60 (27) 0.5175	0/60 (28) NC	1/60 (28) 0.5091	0/60 (31) NC
Pancreas	B-Adenoma, Acinar Cell	1/60 (28) 0.5815	0/60 (28) 0.5000	3/60 (29) 0.3191	0/60 (31) 0.5254
	B-Adenoma, Islet Cell	0/60 (27) 0.1642	1/60 (28) 0.5091	2/60 (28) 0.2545	2/60 (32) 0.2899
	M-Carcinoma, Islet Cell	1/60 (28) 0.7565	0/60 (28) 0.5000	0/60 (28) 0.5000	0/60 (31) 0.5254
	B-Adenoma, Islet Cell/ M-Carcinoma, Islet Cell	1/60 (28) 0.3321	1/60 (28) NC	2/60 (28) 0.5000	2/60 (32) 0.5508

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

**Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Parathyroid	B-Adenoma	0/58 (26) 0.3415	1/57 (27) 0.5094	0/60 (28) NC	1/57 (30) 0.5357
Pituitary	B-Adenoma	35/60 (47) 0.9957	36/60 (46) 0.4265	27/60 (41) 0.7421	21/60 (41) 0.9795
	M-Carcinoma	0/60 (27) 0.5175	0/60 (28) NC	1/60 (28) 0.5091	0/60 (31) NC
	B-Adenoma/M-Carcinoma	35/60 (47) 0.9953	36/60 (46) 0.4265	28/60 (41) 0.6573	21/60 (41) 0.9795
	M-Malignant Schwannoma	0/60 (27) 0.5175	0/60 (28) NC	1/60 (28) 0.5091	0/60 (31) NC
Seminal Vesicle	M-Fibrosarcoma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

**Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Skin/Subcutis	B-Adenoma, Sebaceous	0/58 (26) 0.6506	2/59 (28) 0.2642	0/59 (27) NC	0/59 (30) NC
	M-Carcinoma, Sebaceous	0/58 (26) 0.5312	0/59 (27) NC	2/59 (28) 0.2642	0/59 (30) NC
	B-Adenoma, Sebaceous/ M-Carcinoma, Sebaceous	0/58 (26) 0.5885	2/59 (28) 0.2642	2/59 (28) 0.2642	0/59 (30) NC
	B-Basal Cell Tumor	0/58 (26) 0.5265	1/59 (28) 0.5185	1/59 (28) 0.5185	0/59 (30) NC
	B-Fibroma	2/58 (27) 0.8929	1/59 (28) 0.5139	1/59 (27) 0.5000	0/59 (30) 0.7801
	M-Fibrosarcoma	0/58 (26) 0.5269	5/59 (31) 0.0406 \$	0/59 (27) NC	2/59 (32) 0.3001
	B-Fibroma/M-Fibrosarcoma	2/58 (27) 0.7612	6/59 (32) 0.1892	1/59 (27) 0.5000	2/59 (32) 0.3733
	B-Hemangiopericytoma	0/58 (26) 0.5135	1/59 (28) 0.5185	0/59 (27) NC	0/59 (30) NC
	B-Keratoacanthoma	1/58 (27) 0.5828	0/59 (27) 0.5000	3/59 (28) 0.3194	0/59 (30) 0.5263
	B-Papilloma, Squamous Cell	1/58 (27) 0.8219	1/59 (28) 0.2545	0/59 (27) 0.5000	0/59 (30) 0.5263
	M-Carcinoma, Squamous Cell	1/58 (27) 0.5352	0/59 (27) 0.5000	0/59 (27) 0.5000	1/59 (31) 0.2813
	B-Keratoacanthoma/ M-Carcinoma, Squamous Cell/ B-Papilloma, Squamous Cell	3/58 (27) 0.7681	1/59 (28) 0.7084	3/59 (28) 0.3524	1/59 (31) 0.7449
	B-Lipoma	2/58 (27) 0.9425	0/59 (27) 0.7547	0/59 (27) 0.7547	0/59 (30) 0.7801
	B-Neural Crest Tumor	1/58 (27) 0.7568	0/59 (27) 0.5000	0/59 (27) 0.5000	0/59 (30) 0.5263
	B-Trichoepithelioma	0/58 (26) 0.4450	1/59 (28) 0.5185	4/59 (29) 0.0696	0/59 (30) NC

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

NC = Not calculable.

**Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Testis	B-Hemangioma	1/60 (28) 0.6455	0/60 (28) 0.5000	1/60 (28) NC	0/60 (31) 0.5254
	B-Interstitial Cell Tumor	2/60 (28) 0.9160	2/60 (29) 0.3191	1/60 (28) 0.5000	0/60 (31) 0.7791
	M-Malignant Mesothelioma	0/60 (27) 0.2719	0/60 (28) NC	0/60 (28) NC	1/60 (31) 0.5345
Thymus	B-Thymoma	1/58 (27) 0.7611	0/60 (28) 0.5091	0/59 (27) 0.5000	0/60 (31) 0.5345
Thyroid	B-Adenoma, C-Cell	8/60 (31) 0.9588	7/59 (31) 0.5000	5/60 (30) 0.7109	3/60 (32) 0.9178
	M-Carcinoma, C-Cell	1/60 (28) 0.7089	1/59 (28) NC	2/60 (28) 0.5000	0/60 (31) 0.5254
	B-Adenoma, C-Cell/ M-Carcinoma, C-Cell	9/60 (31) 0.9691	7/59 (31) 0.6138	7/60 (31) 0.6138	3/60 (32) 0.9532
	B-Adenoma, Follicular Cell	4/60 (30) 0.9013	1/59 (28) 0.8014	6/60 (30) 0.3653	0/60 (31) 0.9475
	M-Carcinoma, Follicular Cell	1/60 (28) 0.5259	0/59 (28) 0.5000	0/60 (28) 0.5000	1/60 (31) 0.2718
	B-Adenoma, Follicular Cell/ M-Carcinoma, Follicular Cell	5/60 (30) 0.8651	1/59 (28) 0.8867	6/60 (30) 0.5000	1/60 (31) 0.9097
	Zymbal Gland	M-Carcinoma	0/60 (27) 0.5175	0/60 (28) NC	1/60 (28) 0.5091

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Adrenal, Cortex	B-Adenoma	1/60 (28) 0.2521	0/60 (27) 0.4909	1/60 (25) 0.7257	2/60 (33) 0.5624
	M-Carcinoma	0/60 (28) 0.5089	0/60 (27) NC	1/60 (25) 0.4717	0/60 (32) NC
	B-Adenoma/M-Carcinoma	1/60 (28) 0.2545	0/60 (27) 0.4909	2/60 (25) 0.4568	2/60 (33) 0.5624
Adrenal, Medulla	B-Pheochromocytoma	0/59 (27) 0.0827	0/60 (27) NC	0/59 (24) NC	2/59 (32) 0.2899
	M-Malignant Pheochromocytoma	1/59 (27) 0.7545	0/60 (27) 0.5000	0/59 (24) 0.4706	0/59 (32) 0.5424
	B-Pheochromocytoma/ M-Malignant Pheochromocytoma	1/59 (27) 0.2432	0/60 (27) 0.5000	0/59 (24) 0.4706	2/59 (32) 0.5645
Brain	B-Mixed Glioma	0/60 (28) 0.5133	0/60 (27) NC	1/60 (26) 0.4815	0/60 (32) NC
	M-Glioma	1/60 (28) 0.7500	0/60 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
	B-Mixed Glioma/M-Glioma	1/60 (28) 0.6503	0/60 (27) 0.4909	1/60 (26) 0.7358	0/60 (32) 0.5333
	M-Malignant Oligodendroglio*	0/60 (28) 0.5089	0/60 (27) NC	1/60 (25) 0.4717	0/60 (32) NC
Cervix	B-Polyp, Stromal	0/60 (28) 0.0834	0/60 (27) NC	0/60 (25) NC	2/60 (33) 0.2885
	M-Sarcoma, Stromal	1/60 (28) 0.7500	0/60 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
	B-Polyp, Stromal/ M-Sarcoma, Stromal	1/60 (28) 0.2460	0/60 (27) 0.4909	0/60 (25) 0.4717	2/60 (33) 0.5624
Clitoral Gland	M-Carcinoma	1/60 (28) 0.7500	0/60 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
Hemolympho- Reticular System	M-Histiocytic Sarcoma	0/60 (28) 0.5089	1/60 (27) 0.4909	0/60 (25) NC	0/60 (32) NC
	M-Malignant Lymphoma	1/60 (29) 0.8077	2/60 (28) 0.4866	1/60 (25) 0.7163	0/60 (32) 0.5246

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

NC = Not calculable.

**Table 2B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Kidney	B-Adenoma, Tubule Cell, Amp*	1/60 (28) 0.7591	1/60 (28) NC	1/60 (26) 0.7358	0/60 (32) 0.5333
	M-Carcinoma, Tubule Cell, Amp*	1/60 (28) 0.4706	2/60 (28) 0.5000	0/60 (25) 0.4717	2/60 (34) 0.5736
	B-Adenoma, Tubule Cell, Amp/ M-Carcinoma, Tubule Cell Amp*	1/60 (28) 0.4573	2/60 (28) 0.5000	1/60 (26) 0.7358	2/60 (34) 0.5736
	B-Lipoma	0/60 (28) 0.2920	0/60 (27) NC	0/60 (25) NC	1/60 (33) 0.5410
Liver	B-Adenoma, Hepatocellular	0/60 (28) 0.5089	0/60 (27) NC	1/60 (25) 0.4717	0/60 (32) NC
Mammary Gland	B-Adenoma	1/59 (28) 0.0394*	0/59 (27) 0.4909	0/60 (25) 0.4717	4/60 (34) 0.2437
	M-Carcinoma	17/59 (36) 0.3157	23/59 (40) 0.2528	19/60 (35) 0.3604	24/60 (43) 0.2964
	B-Adenoma/M-Carcinoma	18/59 (36) 0.2689	23/59 (40) 0.3357	19/60 (35) 0.4508	26/60 (44) 0.2785
	B-Fibroadenoma	21/59 (38) 0.1913	23/59 (37) 0.3551	23/60 (35) 0.2510	29/60 (44) 0.2241
Ovary	B-Granulosa Theca Cell Tumor	0/60 (28) 0.5089	1/60 (27) 0.4909	0/60 (25) NC	0/59 (32) NC
	B-Hemangioma	0/60 (28) 0.6460	2/60 (28) 0.2455	0/60 (25) NC	0/59 (32) NC
	M-Sertoli Cell Tumor, Malig*	0/60 (28) 0.5044	1/60 (28) 0.5000	0/60 (25) NC	0/59 (32) NC
Pancreas	B-Adenoma, Acinar Cell	0/60 (28) 0.5044	1/60 (28) 0.5000	0/60 (25) NC	0/59 (32) NC
	B-Adenoma, Islet Cell	1/60 (28) 0.6485	0/60 (27) 0.4909	1/60 (25) 0.7257	0/59 (32) 0.5333
	M-Carcinoma, Islet Cell	0/60 (28) 0.5089	1/60 (27) 0.4909	0/60 (25) NC	0/59 (32) NC
	B-Adenoma, Islet Cell/ M-Carcinoma, Islet Cell	1/60 (28) 0.7605	1/60 (27) 0.7455	1/60 (25) 0.7257	0/59 (32) 0.5333

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

NC = Not calculable.

**Table 2B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Pituitary	B-Adenoma	47/59 (54) 0.8117	52/60 (56) 0.2428	51/60 (54) 0.1599	43/60 (52) 0.6380
	M-Carcinoma	1/59 (27) 0.5068	0/60 (27) 0.5000	1/60 (25) 0.7353	1/60 (33) 0.2983
	B-Adenoma/M-Carcinoma	48/59 (54) 0.8647	52/60 (56) 0.3480	52/60 (54) 0.1351	44/60 (53) 0.7239
Skin/Subcutis	B-Keratoacanthoma	0/60 (28) 0.5044	1/59 (28) 0.5000	0/60 (25) NC	0/60 (32) NC
	B-Papilloma, Squamous Cell	1/60 (28) 0.7500	0/59 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
	B-Keratoacanthoma/ B-Papilloma, Squamous Cell	1/60 (28) 0.8164	1/59 (28) NC	0/60 (25) 0.4717	0/60 (32) 0.5333
	M-Carcinoma, Trichoepitheli*	1/60 (29) 0.7434	0/59 (27) 0.4821	0/60 (25) 0.4630	0/60 (32) 0.5246
	M-Fibrosarcoma	1/60 (28) 0.4991	0/59 (27) 0.4909	1/60 (25) 0.7257	1/60 (33) 0.2885
	M-Sarcoma	1/60 (28) 0.7500	0/59 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
	M-Sarcoma	0/60 (28) 0.5044	1/60 (28) 0.5000	0/60 (25) NC	0/60 (32) NC
Thoracic Cavity	B-Hibernoma	0/60 (28) 0.2920	0/60 (27) NC	0/60 (25) NC	1/60 (33) 0.5410
Thyroid	B-Adenoma, C-Cell	2/60 (29) 0.1984	5/60 (30) 0.2260	2/59 (25) 0.6360	6/60 (35) 0.1983
	B-Adenoma, Follicular Cell	1/60 (28) 0.8142	2/60 (28) 0.5000	1/59 (25) 0.7257	0/60 (32) 0.5333
	M-Carcinoma, Follicular Cell	1/60 (29) 0.7434	0/60 (27) 0.4821	0/59 (25) 0.4630	0/60 (32) 0.5246
	B-Adenoma, Follicular Cell/ M-Carcinoma, Follicular	2/60 (29) 0.9119	2/60 (28) 0.6809	1/59 (25) 0.4435	0/60 (32) 0.7781
Urinary Bladder	B-Papilloma, Transitional C*	0/60 (28) 0.2857	0/60 (27) NC	0/60 (25) NC	1/59 (32) 0.5333
Uterus	B-Polyp, Endometrial Stromal	0/60 (28) 0.1265	2/60 (28) 0.2455	2/60 (25) 0.2177	3/60 (33) 0.1516

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

NC = Not calculable.

**Table 2B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats
(Continued)**

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	5 mg Low (L) P - L vs. C	10 mg Mid (M) P - M vs. C	20 mg High (H) P - H vs. C
Vagina	B-Fibroma	1/60 (28) 0.7500	0/60 (27) 0.4909	0/60 (25) 0.4717	0/60 (32) 0.5333
	B-Papilloma, Squamous Cell	0/60 (28) 0.2920	0/60 (27) NC	0/60 (25) NC	1/60 (33) 0.5410
	M-Sarcoma	0/60 (28) 0.5089	0/60 (27) NC	1/60 (25) 0.4717	0/60 (32) NC
	M-Schwannoma, Malignant	1/60 (29) 0.7434	0/60 (27) 0.4821	0/60 (25) 0.4630	0/60 (32) 0.5246
Zymbal Gland	M-Carcinoma	1/60 (28) 0.6503	0/60 (27) 0.4909	1/60 (26) 0.7358	0/60 (32) 0.5333

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 3A: Intercurrent Mortality Rate in Male Mice

Week / Type of Death	Vehicle Control		Low		Mid		High		Positive Control	
	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %
14 - 27					2	8.00	1	4.00		
Planned intermittent sacrifice									10	
Terminal sacrifice	25	100.00	25	100.00	23	92.00	24	96.00		
Total	25		25		25		25			
Test	All Dose Groups		Vehicle Control vs. Low		Vehicle Control vs. Mid		Vehicle Control vs. High			
Dose-Response (Likelihood Ratio)	0.3511		NC		0.0935		0.2390			
Homogeneity (Log-Rank)	0.2814		NC		0.1531		0.3173			

#All Cum. % Cumulative Percentage except for Terminal sacrifice; NC = Not calculable.

Table 3B: Intercurrent Mortality Rate in Female Mice

Week / Type of Death	Vehicle Control		Low		Mid		High		Positive Control	
	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %
14 - 27							2	8.00		
Planned intermittent sacrifice									10	
Terminal sacrifice	25	100.00	25	100.00	25	100.00	23	92.00		
Total	25		25		25		25			
Test	All Dose Groups		Vehicle Control vs. Low		Vehicle Control vs. Mid		Vehicle Control vs. High			
Dose-Response (Likelihood Ratio)	0.0179*		NC		NC		0.0935			
Homogeneity (Log-Rank)	0.1058		NC		NC		0.1531			

#All Cum. % Cumulative Percentage except for Terminal sacrifice; NC = Not calculable.
* = Significant at 5% level;

Table 4A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Mice

Organ name	Tumor name	Vehicle (VC)	Low (L)	Mid (M)	High (H)	Positive (PC)
		0 mg P - Trend	10 mg P - L vs. VC	30 mg P - M vs. VC	75 mg P - H vs. VC	0 mg P - PC vs. VC
Harderian Glands	Carcinoma	0/25 (25) 0.7475	1/25 (25) 0.5000	0/25 (24) NC	0/25 (25) NC	
Lungs With Bronchi	Alveolar-Bronchiolar Adenoma	1/25 (25) 0.2682	1/25 (25) NC	3/25 (24) 0.2890	2/25 (25) 0.5000	10/10 (10) 0.0000#
Spleen	Hemangiosarcoma	1/25 (25) 0.6455	2/25 (25) 0.5000	0/25 (24) 1.0000	1/25 (25) NC	7/10 (7) 0.0000#
Stomach	Papilloma	0/25 (25) 0.2525	0/25 (25) NC	0/25 (24) NC	1/25 (25) 0.5000	
Testes	Hemangiosarcoma	0/25 (25) 0.7475	1/25 (25) 0.5000	0/25 (24) NC	0/25 (25) NC	
Thyroid Glands	Follicular Cell Adenoma	1/25 (25) 0.7475	0/25 (25) 1.0000	1/25 (24) 0.7449	0/25 (25) 1.0000	
Whole body	Hemangiosarcoma	1/24 (25) 0.7353	3/22 (25) 0.3046	0/24 (25) 1.0000	1/24 (25) NC	7/0 (10) 0.0000#

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 4B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Mice

Organ name	Tumor name	Vehicle (VC)	Low (L)	Mid (M)	High (H)	Positive (PC)
		0 mg P - Trend	10 mg P - L vs. VC	30 mg P - M vs. VC	75 mg P - H vs. VC	0 mg P - PC vs. VC
Cavity, Nasal	Adenocarcinoma	0/25 (25) 0.2424	0/25 (25) NC	0/25 (25) NC	1/25 (24) 0.4898	
Harderian Glands	Adenoma	0/25 (25) 0.6186	1/25 (25) 0.5000	1/25 (25) 0.5000	0/25 (24) NC	
	Carcinoma	0/25 (25) 0.8093	2/25 (25) 0.2449	0/25 (25) NC	0/25 (24) NC	
	Adenoma/Carcinoma	0/25 (25) 0.7847	3/25 (25) 0.1173	1/25 (25) 0.5000	0/25 (24) NC	
Lungs With Bronchi	Alveolar-Bronchiolar Adenoma	3/25 (25) 0.7242	0/25 (25) 1.0000	1/25 (25) 0.9451	1/25 (24) 0.9403	10/10 (10) 0.0000#
Multicentric	Sarcoma	0/25 (25) 0.4949	0/25 (25) NC	1/25 (25) 0.5000	0/25 (24) NC	0/10 (1) NC
Ovaries	Hemangiosarcoma	1/25 (25) 1.0000	0/25 (25) 1.0000	0/25 (25) 1.0000	0/25 (24) 1.0000	
Spleen	Hemangiosarcoma	1/25 (25) 0.6798	0/25 (25) 1.0000	2/25 (25) 0.5000	0/25 (24) 1.0000	5/10 (5) 0.0000#
Whole body	Hemangiosarcoma	2/23 (25) 0.8239	0/25 (25) 1.0000	2/23 (25) NC	0/24 (25) 1.0000	5/0 (10) 0.0001#

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;

NC = Not calculable.

Figure 1A: Kaplan-Meier Survival Functions for Male Rats

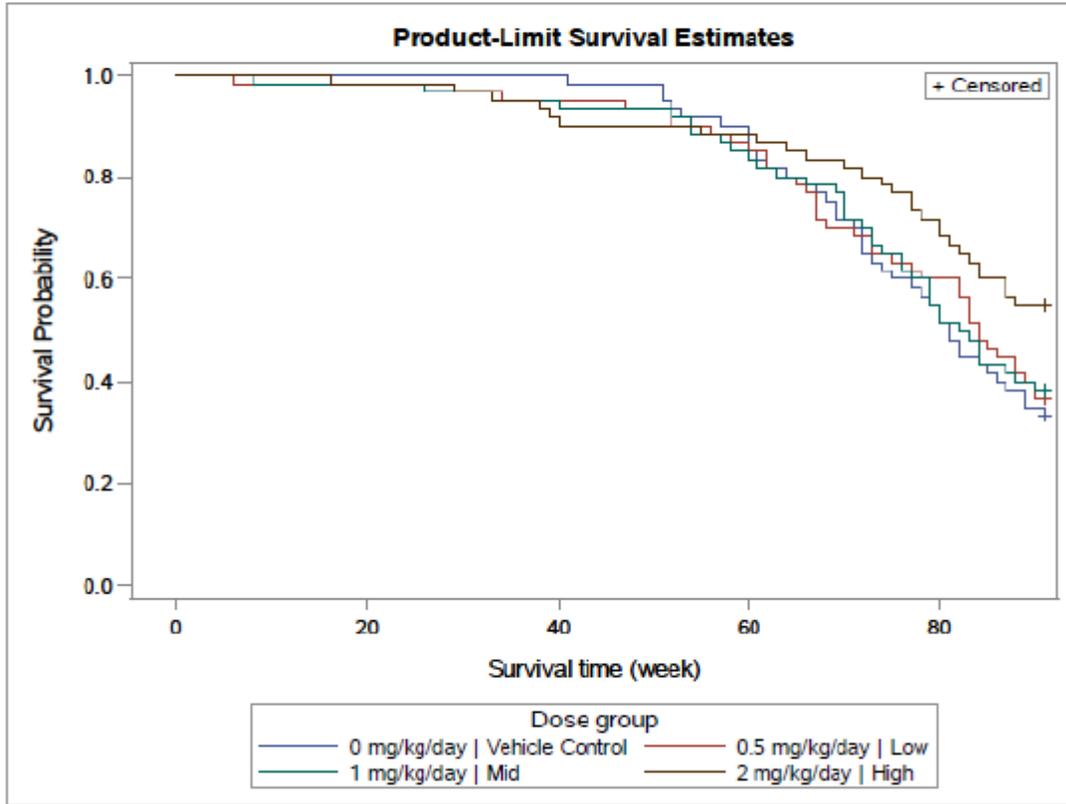


Figure 1B: Kaplan-Meier Survival Functions for Female Rats

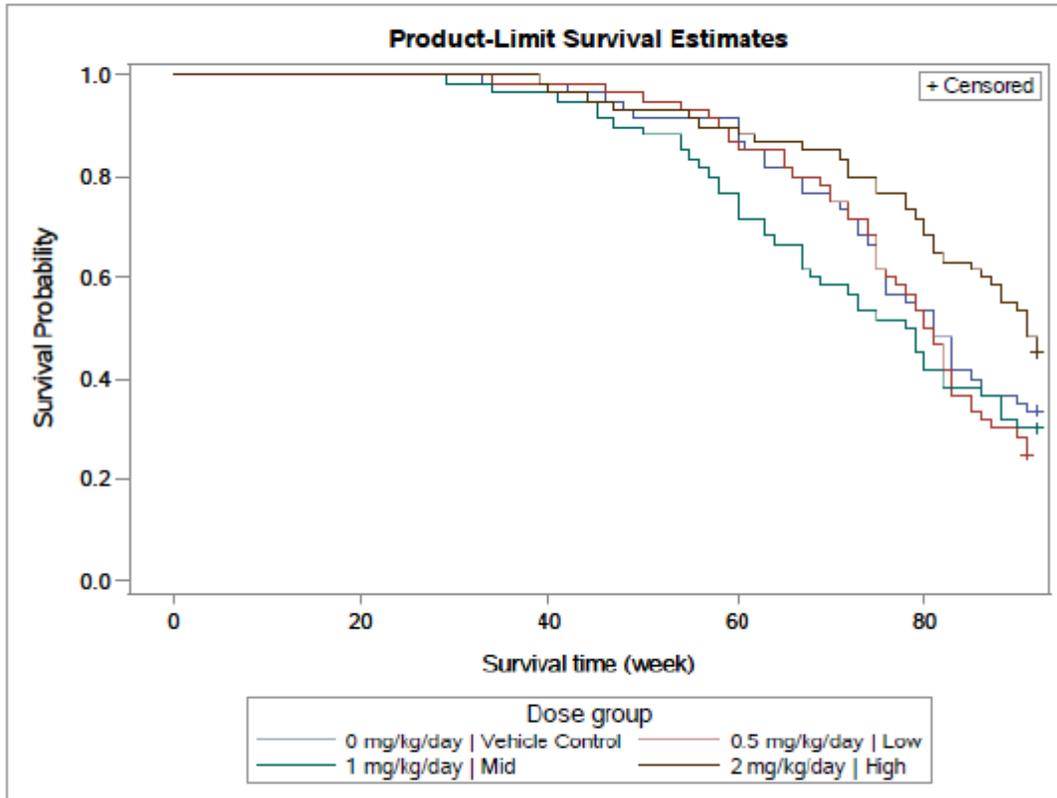


Figure 2A: Kaplan-Meier Survival Functions for Male Mice

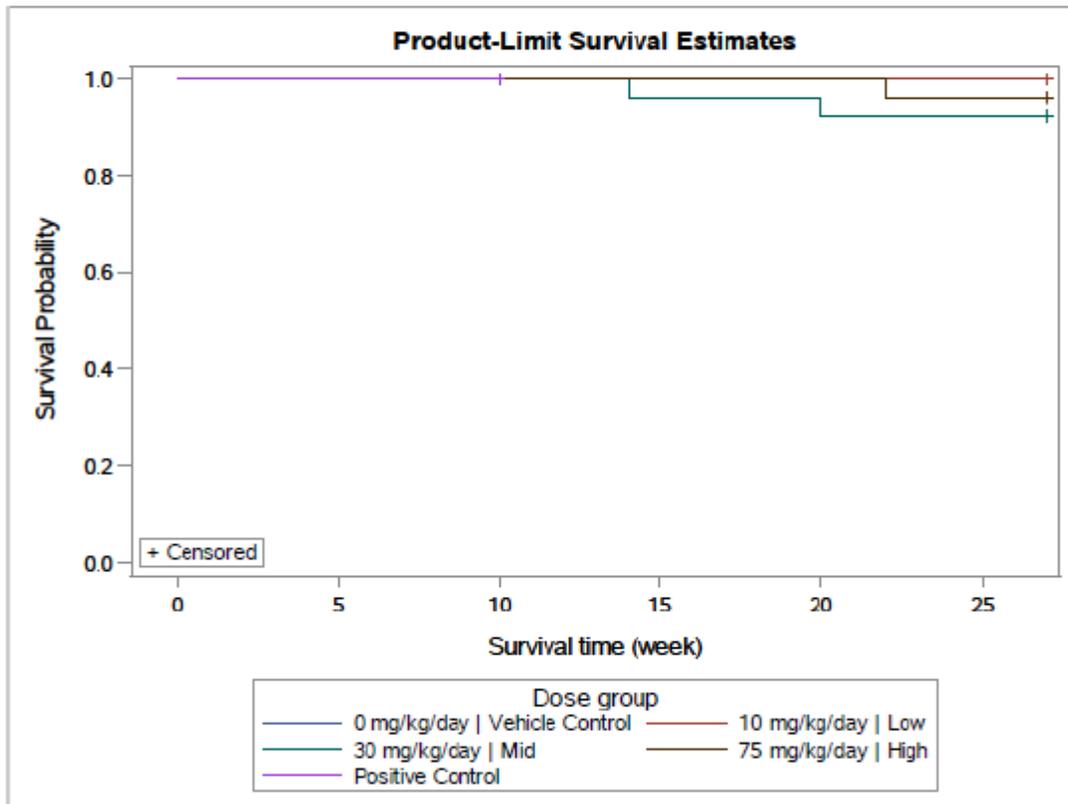
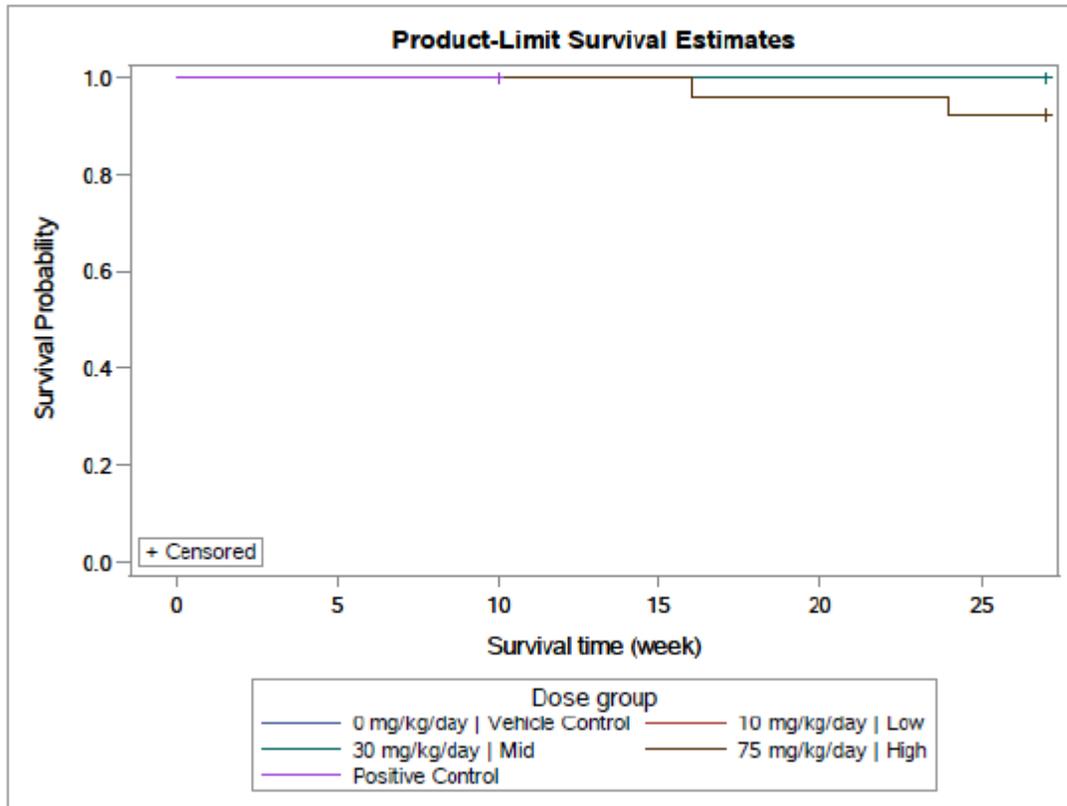


Figure 2B: Kaplan-Meier Survival Functions for Female Mice



6. References

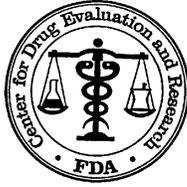
- Bailer, A.J, Portier, C.J. (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples." *Biometrics*, 44, 417-431.
- Bieler, G.S. and Williams, R.L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.
- Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). May 2001.
- Haseman, J. (1983). "A re-examination of false-positive rates for carcinogenesis studies", *Fundamental and Applied Toxicology*, 3: 334-339.
- Lin KK (2000) Carcinogenicity Studies of Pharmaceuticals. In: *Encyclopedia of Biopharmaceutical Statistics*, ed. Shein-Chung Chow, Marcel Dekker, New York.
- Lin K.K. and Rahman A.M. (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.
- Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf. (1980) "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva*, 311-426.
- Rahman, A.M., 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*, 18:5, 849-858.
- Rahman, A.M., and Lin, K.K. (2009), "Design and Analysis of Chronic Carcinogenicity Studies of Pharmaceuticals in Rodents", in "Design and Analysis of Clinical Trials with Time-to-Event Endpoints", K.E Peace, Editor, Chapman & Hall/CRC, Taylor & Francis Group, LLC, Boca Raton, FL, London, and New York.

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

/s/

DARREN B FEGLEY
03/06/2017

AISAR H ATRAKCHI
03/07/2017



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 #: 209,241

Drug Name: INGREZZA (valbenazine tosylate) 40 mg capsules

Indications: Treatment of Tardive Dyskinesia

Applicant: Neurocrine Biosciences

Dates: Submitted: 08/11/2016
PDUFA due date: 04/11/2017

Review Priority: Priority

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Thomas Birkner, Ph.D.

Concurring Reviewers: Peiling Yang, Ph.D.
H.M. James Hung, Ph.D.

Medical Division: Division of Psychiatry Products

Clinical Team: Michael Davis, MD; Javier Muniz, MD (TL); Brian Miller, MD

Project Manager: Jasmeet (Mona) Kalsi

Keywords: mixed models, multiple endpoints, tipping point

Table of Contents

1	EXECUTIVE SUMMARY	5
2	INTRODUCTION	6
2.1	OVERVIEW.....	6
2.2	DATA SOURCES	8
3	STATISTICAL EVALUATION	8
3.1	DATA AND ANALYSIS QUALITY	8
3.2	EVALUATION OF EFFICACY	9
3.2.1	<i>Study Design and Endpoints</i>	9
3.2.2	<i>Statistical Methodologies</i>	10
3.2.3	<i>Patient Disposition, Demographic and Baseline Characteristics</i>	14
3.2.4	<i>Results and Conclusions</i>	18
3.3	EVALUATION OF SAFETY	36
4	FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	36
4.1	GENDER, RACE, AGE, AND GEOGRAPHIC REGION	36
4.2	OTHER SPECIAL/SUBGROUP POPULATIONS	38
5	SUMMARY AND CONCLUSIONS	42
5.1	STATISTICAL ISSUES	42
5.2	COLLECTIVE EVIDENCE.....	43
5.3	CONCLUSIONS AND RECOMMENDATIONS	43
5.4	LABELING RECOMMENDATIONS	43
	APPENDICES.....	44

LIST OF TABLES

Table 1. List Of All Studies Included In Review.....	7
Table 2. Study 1202 Summary of Analysis Sets.....	14
Table 3. Study 1202 Subject Enrollment and Disposition (All Randomized Subjects).....	15
Table 4. Study 1202 Select Demographics and Baseline Characteristics (Safety Analysis Set)	15
Table 5. Study 1304 Summary of Analysis Sets.....	16
Table 6. Study 1304 Subject Enrollment and Disposition Through Week 6 (All Randomized Subjects at Day -1)	17
Table 7. Study 1304 Select Demographics and Baseline Characteristics (Safety Analysis Set)	17
Table 8. Study 1202 Summary of Dosing Titration by Visit and Treatment (Safety Analysis Set)	18
Table 9. Study 1202 Primary Endpoint Results (Per-Protocol)	19
Table 10. Study 1202 Primary Endpoint Results (ITT)	19
Table 11. Study 1202 CGI TD Assessment (Week 6) - ITT.....	20
Table 12. Study 1202 Secondary Endpoint Results	20
Table 13. Study 1202 Secondary Endpoint Results (on site raters and ANCOVA)	21
Table 14. Study 1202 Secondary Endpoint Results (on site raters and MMRM).....	21
Table 15. Study 1304 Primary Endpoint Results	22
Table 16. Study 1304 Key Secondary Endpoint Results	23
Table 17. Study 1304 Summary Table of Efficacy Results given Multiplicity Control Procedure.....	23
Table 18. Study 1304 Open-label Extension: Subject Enrollment and Disposition	25
Table 19. Study 1304 Observed Mean AIMS Total Score Change from Baseline During the Extension and Posttreatment Periods (ITT).....	25
Table 20. Study 1202 Primary Endpoint Results by Week 6 Dose	29
Table 21. Study 1202 Sensitivity Analysis Including ‘Randomized but Excluded’ Patients	31
Table 22. Study 1304 Exploratory Subgroup Analysis by Gender	36
Table 23. Study 1304 Exploratory Subgroup Analysis by Race.....	37
Table 24. Study 1304 Exploratory Subgroup Analysis by Age	38
Table 25: Study 1304 Metabolizer Subgroup Frequencies	39
Table 26. Study 1304 Exploratory Subgroup Analysis by CYP2D6 Metabolizer Category	39
Table 27. Study 1304 Exploratory Subgroup Analysis by Disease Category.....	40
Table 28. Study 1304 Exploratory Subgroup Analysis by Antipsychotic Medication Use	41

LIST OF FIGURES

Figure 1. Study 1304 AIMS Mean Score Change from Baseline During the Extension and Posttreatment Periods by Treatment (Arithmetic Mean \pm SEM, ITT).....	26
Figure 2. Study 1202 Frequency of Subjects with Specified Magnitude in AIMS CFB at Week 6.....	27
Figure 3. Study 1304 Frequency of Subjects with Specified Magnitude in AIMS CFB at Week 6 (ITT) .	32
Figure 4: Study 1304 Percent of Patients with Specified Magnitude of AIMS Total Score Improvement at Week 6 (ITT)	33
Figure 5. Study 1304 AIMS Component Mean Profiles (Placebo)	34
Figure 6. Study 1304 AIMS Component Mean Profiles (Valbenazine 40 mg)	35
Figure 7. Study 1304 AIMS Component Mean Profiles (Valbenazine 80 mg)	35
Figure 8. Study 1304 Forest Plot Exploratory Subgroup Analysis of AIMS CFB (80 mg)	42

1 EXECUTIVE SUMMARY

The two positive efficacy studies submitted under this NDA support the claim that the 80 mg dose of Valbenazine reduces the symptoms of Tardive Dyskinesia (TD) as measured by the Abnormal Involuntary Movement Scale (AIMS) in an acute setting (double-blind treatment phase of 6 weeks). The 40 mg dose of Valbenazine appears to be efficacious for some patients with TD, but does not meet the strict evidentiary standard for approval (e.g., no replication of finding).

The two efficacy studies supporting the claim are the Phase 2 study NBI-98854-1202 and the Phase 3 study NBI-98854-1304. Study 1304 (n=225) tested two fixed doses of Valbenazine versus placebo. Patients treated with the 80 mg dose of Valbenazine had an average of -3.1 points greater reduction on the primary efficacy endpoint (i.e., AIMS total score change from baseline to week 6) compared to placebo (95% CI: -4.2, -2.0; $p < 0.0001$). The 40 mg Valbenazine group also had favorable numeric results (average improvement on the AIMS of -1.8 points over placebo). However, the results for the 40 mg group were not statistically significant because the 80 mg dose was not statistically significant ($p=0.056$) on the key secondary endpoint Clinical Global Impression of Change-TD, which was placed before any 40 mg dose tests in the fixed testing sequence.

Patients treated with Valbenazine (dose titration from 25, to 50, to 75 mg every two weeks based on therapeutic response and tolerability) in Study 1202 (n=89) achieved on average a -2.4 point greater reduction in the AIMS total score at the end of week 6 compared to patients treated with placebo (95% CI: -3.7, -1.1). The majority of Valbenazine patients (69%) were titrated to the 75 mg dose by the end of the study and the mean dose at week 6 was 64.4 mg/day. Those findings allow viewing the results from Study 1202 to serve as replication for the Phase 3 results for the 80 mg dose.

This reviewer did not find any major statistical issues for Study 1304, besides the unfortunate order of endpoints in the fixed testing sequence which the Division had recommended revising during the IND stage. The sponsor had decided not to modify their approach.

The review of Study 1202 revealed a number of statistical concerns, which could be partially alleviated by this reviewer's analyses. One major issue was the choice of the per-protocol set as primary efficacy analysis set. Fortunately for the sponsor the efficacy conclusions did not differ when considering the ITT set. However, the 'ITT' set is not issue-free either, as it was turned into a completer set by the choice of primary analysis method (i.e., ANCOVA change from baseline to week 6) implemented with protocol amendment 2. This amendment did not receive full regulatory scrutiny (i.e., no statistical review), because Study 1202 was regarded as non-pivotal. This reviewer explored the efficacy trajectories of the 12 patients who discontinued the study during the double-blind phase (and were excluded from the 'ITT' set). The available AIMS data (prior drop-out) and an exploratory tipping point analysis suggest that omitting those patients did not materially impact the efficacy conclusion of Study 1202.

2 INTRODUCTION

2.1 Overview

2.1.1. Class and Indication

NBI-98854 (Valbenazine) is an orally active vesicular monoamine transporter 2 inhibitor under development by Neurocrine Biosciences for the treatment of Tardive Dyskinesia (TD) (b) (4). TD is a neurological condition characterized by involuntary movements and is associated with long-term neuroleptic drug use.

2.1.2. History of Drug Development

The first Phase 2 studies (e.g., Study 1201) in the sponsor's TD development program (IND 111,591) did not achieve statistically significant efficacy results. Subsequently the sponsor convened a panel of experts, who recommended the use of central raters instead of site raters and a change in the score descriptors for the Abnormal Involuntary Movement Scale (AIMS). Those changes were implemented in Studies 1202 (Phase 2) and 1304 (Phase 3). This IND received Breakthrough Therapy Designation.

Advice to Applicant

Study 1202

The protocol and protocol amendments for Study NBI-98854-1202 appear not to have been reviewed by a statistical reviewer, since this Phase 2 study was not considered a pivotal trial. Below is the clinical reviewer's view of Amendment 2 to the protocol (Review noted 04/18/2014):

“The protocol was amended to include scoring of AIMS items 1-7 by a central, blinded video rater, to make the AIMS dyskinesia total score (items 1-7) change from baseline derived from the blinded, central video rater's scores the primary endpoint of the study and to make the AIMS dyskinesia total score as scored by the site's certified, independent AIMS rater a secondary efficacy endpoint. **As previously discussed with the sponsor, this study is not a pivotal trial; therefore, these changes are deemed acceptable.** [...] No action indicated from a clinical perspective.” [bold highlight added]

At the June 24, 2014 EOP2 meeting the sponsor inquired whether a single pivotal study would suffice to confirm efficacy. The Division disagreed. However, after the sponsor submits the draft CSR of Study 1202 for a preliminary review, the Division states in a post meeting note “... we believe that study 1202 could potentially provide evidence in support of NBI-98854 registration for the treatment of TD.”

Study 1304

The sponsor submitted the protocol (and amendments) and the SAP (and revisions) for FDA review. The final SAP was submitted on September 10, 2015 prior to unblinding of the study. Statistical comments regarding the testing sequence given two dose groups and the primary and key secondary endpoints were communicated. The Division recommended testing the primary endpoint for both high and low dose first, before testing the key secondary endpoint. The sponsor did not implement that recommendation and proceeded with testing the primary and key secondary endpoint first for the high dose, before testing any endpoint for the lower dose. Further comments were conveyed regarding sensitivity analyses, choice of key secondary endpoints and specifications of the primary statistical model.

2.1.3. Specific Studies Reviewed

The studies selected for review are the Phase 2 Study NBI-98854-1202 and the Phase 3 study NBI-98854-1304. Both studies are presented as the pivotal studies supporting the TD claim by the sponsor. Table 1 provides some basic characteristics for both studies.

Table 1. List Of All Studies Included In Review

Study Number	Phase and Design	Treatment Period	# of Subjects per Arm	Study Population
1202	MC, R, DB, PG, PC trial	6 weeks	NBI-98854 25-75mg/ 51 Placebo/ 51	Tardive Dyskinesia
1304	MC, R, DB, PG, PC trial	6 weeks	NBI-98854 40mg/ 76 80mg/ 80 Placebo/ 78	Tardive Dyskinesia

* MC: multi-center, R: randomized, DB: double-blind, PG: parallel group, PC: placebo controlled

2.1.4. Major Statistical Issues

Major statistical issues include the choice of the per-protocol set as primary analysis set and the definition of the 'ITT' set leading to a completer set in Study 1202. Not per se a statistical issue, but the pre-specified fixed testing procedure in Study 1304 precludes the formal statistical testing of the primary efficacy endpoint for the 40 mg dose (nominal p-value of 0.0021).

2.2 Data Sources

The review encompasses the protocols, statistical analysis plans, and study reports for Studies 1202 and 1304. Also considered were the Integrated Summary of Efficacy and the 120-Day Safety Update submission for Study 1304.

Most analyses performed by this reviewer are based on the following datasets.

Study 1202:

Primary efficacy dataset: a_aimscr

Efficacy based on independent site raters: a_aims

Secondary efficacy: cgitd

Study 1304:

Primary and key secondary efficacy: A_aims; cgitd

Exploratory efficacy analysis (AIMS components): qs

The electronic location of the initial submission is: <\\Cdsub1\evsprod\NDA209241\0002>

The 120-Day Safety Update can be accessed at the following location:

<\\Cdsub1\evsprod\NDA209241\0014>

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Some issues were encountered regarding the completeness of the analysis dataset a_aimscr for Study 1202. The psychiatric diagnosis variable 'diag' was not included with the primary efficacy dataset, although it is a predictor in the primary efficacy analysis model. This reviewer found the 'diag' variable in the medical history dataset 'mhdiag'. Furthermore the opening of a number of SAS datasets for Study 1304 relied on writing formatting statement code in SAS to overcome the issue of missing formats. This reviewer was able to replicate the AIMS Dyskinesia Total Score data based on a dataset containing the seven components of the AIMS Dyskinesia Total Score. The randomization process for both studies (1202 and 1304) appears valid (see appendix for Figures A1 and A2, which display the assignment of patients over time to the treatment groups). The SAP for Study 1202 (dated 10/10/2013) was submitted with the NDA. It is unclear whether the SAP was ever submitted to the IND. If the SAP was indeed submitted it seems likely that it was not reviewed by the Biometrics Division because Study 1202 was regarded as exploratory. The SAP for Study 1304 was submitted to the IND and reviewed. The Division provided comments; most of which the sponsor accepted.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

Study NBI-98854-1202

Study 1202 was conducted between February and December 2013 at 29 study centers in the US and Puerto Rico. It is a Phase 2, randomized, double-blind, dose-titration, placebo-controlled study to evaluate the efficacy, safety, and tolerability of NBI-98854 (25 mg titrated to 50 mg and subsequently to 75 mg once daily for 6 weeks) in subjects with schizophrenia or schizoaffective disorder, or mood disorder with TD, or GI disorder with TD.

Efficacy Endpoints

The primary efficacy endpoint was the change from baseline in the Abnormal Involuntary Movement Scale (AIMS) dyskinesia total score (sum of Items 1-7) based on the blinded, central AIMS video raters' assessment.

The AIMS total dyskinesia score calculated from the site's certified, independent AIMS rater's assessment and the Clinical Global Impression of Tardive Dyskinesia (CGI-TD) were evaluated as secondary efficacy endpoints.

Dose titration algorithm

The starting dose for the patients randomized to Valbenazine was 25 mg once daily. Patients were eligible for a dose escalation at the end of week 2 (to 50 mg) and then again at the end of week 4 (to 75 mg). Dose escalation eligibility was based on a 2-part evaluation of the current dose: (1) if the subject had a score of mild, moderate, or severe on any of the AIMS items 1 to 7 as assessed by the independent on-site AIMS rater; and (2) if the physician investigator determined that a dose escalation was acceptable based on the safety and tolerability of the current dose. At any time after week 2, the physician investigator could decrease the dose to the previous dose for any patient who was unable to tolerate a given dose increase (Study report page 28).

Protocol Amendment

A Scientific Advisory Board recommended the "triple blind" central rating by two movement disorder neurologists (instead of independent site raters as specified in the original protocol), consensus scores for Items 1 through 7, and alternative descriptors for the AIMS scores. The recommendations were implemented with protocol amendment 2 before study completion and database lock. The COA staff found the changes to the score descriptors acceptable.

Study NBI-98854-1304

Study 1304 was conducted between November 2014 and March 2016 at 63 sites in North America and Puerto Rico. It is a Phase 3, double-blind, parallel-group, fixed-dose study to evaluate the efficacy, safety, and tolerability of NBI-98854 (40 mg once daily or 80 mg once daily) compared with placebo (1:1:1) in subjects with schizophrenia or schizoaffective disorder, or mood disorder with TD. The study design includes a double-blind, placebo-controlled treatment period for 6 weeks followed by an NBI-98854 extension period for 42 weeks. During the extension period, all subjects received NBI-98854. The extension period of this study was ongoing at the time of NDA submission; results were included with the 120-Day Safety Update. The sponsor's approach to control the Type I error was a fixed testing sequence (test primary followed by key secondary efficacy endpoint for 80 mg Valbenazine; if statistically significant results then test primary and key secondary endpoints for the 40 mg Valbenazine dose).

Efficacy Endpoints

The primary efficacy endpoint was the AIMS dyskinesia total score mean change from baseline to week 6 based on the blinded, central video raters' assessment.

The key secondary efficacy endpoint was the mean CGI-TD score at week 6.

3.2.2 Statistical Methodologies

Study 1202

Analysis sets

The **Safety analysis set** includes all subjects who are randomized to a treatment group and receive at least one dose of study drug.

The **ITT analysis set** includes all subjects in the safety analysis set who have an evaluable blinded, central video rater's AIMS dyskinesia total score change from baseline (CFB) value at one or more scheduled assessment times during the double-blind treatment period.

Reviewer's note: After the switch to central raters and to ANCOVA as primary analysis and the subsequent central rating only at baseline and week 6 this ITT set is effectively a completer set.

The **Per-Protocol (PP) analysis set** will include subjects in the ITT analysis set who meet the following criteria:

1. An evaluable blinded, central video rater's AIMS dyskinesia total score CFB value at Week 6,
2. For subjects in the Valbenazine group, a quantifiable Valbenazine plasma concentration at Week 6, and
3. No efficacy-related important protocol deviations.

For the week 6 analysis only the first two criteria listed above are applied.

Reviewer's note: The PP analysis set was specified as the primary efficacy analysis set by the sponsor.

Primary analysis for primary endpoint: AIMS total score change from baseline (BL) to Week 6

The primary analysis of the AIMS data is an analysis of covariance (ANCOVA) of the blinded, central video rater total score CFB data at Week 6 using the PP analysis set. The ANCOVA model includes the baseline AIMS total score as a covariate and treatment group and disease category as fixed effects.

Analysis of secondary endpoint: CGI-TD at Week 6

Hypothesis tests comparing the Valbenazine group to placebo at Week 6 were performed for the CGI-TD scores using an analysis of variance (ANOVA) model. The ANOVA model includes treatment group and disease category as fixed effects.

Reviewer's note: The CGI-TD was not clearly pre-specified as key secondary efficacy endpoint in Study 1202. The protocol and SAP list the AIMS (independent site raters) and the CGI-TD as secondary efficacy endpoints. If assuming a fixed sequence testing procedure, it is not clear whether the CGI-TD would be tested before or after the AIMS (independent site raters). Note that the results for the AIMS (independent site raters) were not statistically significant.

Study 1202 protocol, amendments and conduct

- Original protocol [dated September 25, 2012]: The primary efficacy variable was to be assessed by independent site raters (BL, Week 2, 4, 6) and the primary analysis was MMRM.
- Protocol amendment 1 [February 4, 2013]: Minor changes.
- Protocol amendment 2 [September 26, 2013]: The primary efficacy assessment is now to be made by central raters (BL, Week 2, 4, 6) and primary analysis is changed to ANCOVA of CFB to Week 6. MMRM is retained as supplemental analysis.
- SAP [dated October 10, 2013]: Implements changes of protocol amendment 2.

No subjects were randomized into the study under the original protocol. Of the 102 subjects randomized, 62 were randomized under protocol amendment No. 1 and 40 were randomized under amendment No. 2.

Deviations from the protocol and SAP occurred in the conduct of the study as documented in the 1202 Study Report [dated July 7, 2015]: Central Ratings were only conducted at BL and Week 6; and the supplemental MMRM analysis was not done as a result.

“As described in the SAP, it was originally planned to have the central video raters’ score the AIMS videos at all study visits (baseline, Weeks 2, 4, and 6, and follow-up visit at Week 8); however, only the baseline (Day -1) and Week 6 videos were scored due to the large number of video recordings across timepoints and the nature of the study design. As a result, all summaries and analyses of the central video raters’ AIMS dyskinesia total score CFB data (including responder analyses) were performed only for Week 6. The supplemental analysis of these data using the MMRM model was also not conducted. Furthermore, the original ITT analysis set specification included all subjects in the safety analysis set who have an evaluable blinded, central video raters’ AIMS dyskinesia total score CFB value at one or more scheduled assessment times during the double-blind treatment period; however, due to the above modifications, this specification is now equivalent to including subjects with a central video raters’ AIMS dyskinesia total score CFB value at Week 6.” (Study report p. 67)

Note the sponsor’s response to the information request by Michael Davis requesting information why the supplemental MMRM analysis was not performed (sponsor response SN 25 (01/26/2017)): “data generated from this analysis [MMRM] would be uninformative due to the titration design of the study with subjects on different doses (25 mg or 50 mg) at intermediate time points”.

The change in the planned primary analysis from MMRM to ANCOVA with Amendment 2 and the subsequent sponsor decision to employ central raters for BL and Week 6 AIMS assessments only was not discussed between the sponsor and the Division. A comparison of the AIMS central rating results of the original primary analysis method (MMRM) with the final analysis method (ANCOVA) is not feasible, since data based on central ratings is only available for baseline and Week 6. The appropriateness of the ANCOVA in the presence of missing data is based on Missing Completely At Random (MCAR) assumption. The exploration of the impact of missing data is described in the reviewer’s results section of this review.

Study 1304

Analysis Sets

The **safety analysis set** includes all subjects who are randomized to a treatment group and dispensed study drug.

The primary efficacy analysis set is the **ITT analysis set**, which includes all subjects in the safety analysis set who have a baseline (Day-1) AIMS dyskinesia total score value and at least

one post-randomization AIMS dyskinesia total score value reported during the double-blind, placebo-controlled treatment period.

Analysis

The primary efficacy endpoint for this study is the AIMS dyskinesia total score mean change from baseline at Week 6, and the key secondary efficacy endpoint is the CGI-TD mean score at Week 6.

Primary Endpoint: AIMS dyskinesia total score mean change from baseline at Week 6

The primary analysis of this endpoint is a MMRM analysis. The model includes the baseline AIMS dyskinesia total score as a covariate, with treatment group (placebo, Valbenazine 40 mg, or Valbenazine 80 mg), disease category (schizophrenia/schizoaffective disorder or mood disorder), and visit (Week 2, 4, or 6) as fixed effects, and subject as a random effect. The model also includes the following interaction terms: (a) treatment group x visit and (b) baseline x visit.

Secondary: CGI-TD mean score at Week 6

An analysis of the CGI-TD scores at Weeks 2 through 6 was performed using a MMRM model similar to the primary endpoint analysis, but without the covariate (baseline AIMS total score) or the baseline x visit interaction term.

Multiplicity

Fixed sequence: 80 mg AIMS, 80 mg CGI-TD, 40 mg AIMS, 40 mg CGI-TD.

Sensitivity Analyses

A Tipping Point and Jump to Reference Analysis were performed to assess the impact of deviations from the missing at random assumption for missing data under the repeated measures primary analysis model (MMRM).

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

Study 1202

Table 2. Study 1202 Summary of Analysis Sets

Populations	Placebo n (%)	NBI 25-75 mg n (%)	Total n (%)
Randomized	51	51	102
Safety	49 (96)	51 (100)	100 (98)
ITT	44 (86)	45 (88)	89 (87)
PP*	44 (86)	32 (63)	76 (75)

*The Per Protocol Set is included here because it is the pre-specified primary efficacy analysis set.

Of the 102 subjects randomized into Study 1202, 100 (98%) received at least one dose of study drug. Two subjects in the placebo group did not receive study drug. Note that the number of ITT completers is the same as number of subjects included in the ITT population (by definition of ‘ITT’ in this study). The percentage of randomized subjects excluded from the ‘ITT set’ is 14% (n=7) for the placebo and 12% (n=6) for the Valbenazine arm. Patients were excluded from the ‘ITT set’ if they had no week 6 AIMS central rating (Placebo: 10% (n=5), Valbenazine 12% (n=6)), which is for the most part equivalent to early discontinuation due to an adverse event, non-compliance, or withdrawal of consent (Table 3). Note that 11 subjects in the Valbenazine group had no quantifiable drug plasma concentration at week 6. Those subjects were excluded from the PP population.

Table 3. Study 1202 Subject Enrollment and Disposition (All Randomized Subjects)

Number of Subjects	Placebo (N=51) n (%)	NBI-98854 (N=51) n (%)	All Subjects (N=102) n (%)
Received study drug	49 (96.1)	51 (100.0)	100 (98.0)
Discontinued during double-blind treatment period	7 (13.7)	5 (9.8)	12 (11.8)
Completed double-blind treatment period ^a	44 (86.3)	46 (90.2)	90 (88.2)
Discontinued during post-treatment period	0 (0.0)	0 (0.0)	0 (0.0)
Completed study	44 (86.3)	46 (90.2)	90 (88.2)
Reason for Discontinuation			
Adverse event	2 (3.9) ^b	0 (0.0)	2 (2.0)
Protocol deviation	0 (0.0)	0 (0.0)	0 (0.0)
Noncompliance	3 (5.9)	2 (3.9)	5 (4.9)
Withdrawal of consent	1 (2.0)	3 (5.9)	4 (3.9)
Death	0 (0.0)	0 (0.0)	0 (0.0)
Lost to follow-up	1 (2.0)	0 (0.0)	1 (1.0)
Sponsor/Investigator decision	0 (0.0)	0 (0.0)	0 (0.0)

(Source: Study Report p. 69)

Of the 100 subjects in the safety analysis set, 57% were male and 43% female (Table 4). The average age was 56.2 years. The majority of patients were Caucasian (63%), followed by African American (34%).

Table 4. Study 1202 Select Demographics and Baseline Characteristics (Safety Analysis Set)

Variable	Statistic or Category	Placebo (n=49)	NBI 25-75 mg (n=51)	Total (n=100)
Age (years)	Mean	56	57	56
	SD	9.8	10.8	10.3
	Median	57	56	57
	Min, Max	34, 78	32, 78	32, 78
Gender (n[%])	Male	27 (55)	30 (59)	57 (57)
	Female	22 (45)	21 (41)	43 (43)
Race* (n [%])	Black or African American	16 (33)	18 (35)	34 (34)
	Caucasian	30 (61)	33 (65)	63 (63)
Disease Category (n[%])	Schizophrenia/schizoaffective	30 (61)	28 (55)	58 (58)
	Mood Disorder	18 (37)	20 (39)	38 (38)
	Gastrointestinal disorder	1 (2)	3 (6)	4 (4)

Baseline	n	44	45	89
AIMS	Mean	7.9	8.0	8.0
Dyskinesia	SD	4.5	3.5	4.0
Total Score	Median	7	7	7
	Min, Max	1, 23	3, 18	1, 23

(Source: 1202 Study Report p. 73, 75; *omitted races with at most 1 patient per treatment group)

Study 1304

A total of 234 subjects were randomized in the double-blind phase of the study (Table 5). The safety set included 227 of the 234 subjects (two subjects withdrew and returned all study drug and five subjects had no post-baseline safety data collected). The ITT analysis set included 225 subjects (two safety set subjects were excluded because they had no post-baseline AIMS total score).

Table 5. Study 1304 Summary of Analysis Sets

Populations	Placebo	NBI 40 mg	NBI 80 mg	Total
	n (%)	n (%)	n (%)	n (%)
Randomized	78	76	80	234
Safety	76 (97)	72 (95)	79 (99)	227 (97)
ITT	76 (97)	70 (92)	79 (99)	225 (96)

Of the 29 (12.4%) subjects who discontinued during the double-blind treatment period, eight (3.4%) were discontinued because of an adverse event (four NBI 40 mg, two NBI 80 mg, and two placebo subjects). Ten subjects withdrew consent (five NBI 40 mg, four NBI 80 mg, and one placebo subject). Other reasons for discontinuations were lost to follow-up, non-compliance, and sponsor/investigator decision (Table 6).

Table 6. Study 1304 Subject Enrollment and Disposition Through Week 6 (All Randomized Subjects at Day -1)

Number of Subjects	Placebo n (%)	NBI-98854 40 mg n (%)	NBI-98854 80 mg n (%)	All Subjects n (%)
Randomized	78	76	80	234
Completed Placebo-Controlled Period	71 (91.0)	63 (82.9)	71 (88.8)	205 (87.6)
Discontinued Placebo-Controlled Period	7 (9.0)	13 (17.1)	9 (11.3)	29 (12.4)
Reason for discontinuation:				
Adverse event	2 (2.6)	4 (5.3)	2 (2.5)	8 (3.4)
Non-compliance	2 (2.6)	1 (1.3)	0	3 (1.3)
Withdrawal of consent	1 (1.3)	5 (6.6)	4 (5.0)	10 (4.3)
Death	0	0	1 (1.3)	1 (0.4)
Lost to follow-up	2 (2.6)	1 (1.3)	1 (1.3)	4 (1.7)
Sponsor/investigator decision	0	2 (2.6)	1 (1.3)	3 (1.3)
Discontinued at Week 6	2 (2.6)	2 (2.6)	3 (3.8)	7 (3.0)
Reason for discontinuation:				
Adverse event	2 (2.6)	1 (1.3)	1 (1.3)	4 (1.7)
Non-compliance	0	1 (1.3)	1 (1.3)	2 (0.9)
Withdrawal of consent	0	0	1 (1.3)	1 (0.4)

(Source: Study Report p. 70)

Baseline characteristics appear well balanced across treatment groups. The mean age was 56 years; slightly more males (54%) were randomized than females. 56% of subjects were Caucasians, followed by African Americans (38%) (Table 7).

Table 7. Study 1304 Select Demographics and Baseline Characteristics (Safety Analysis Set)

Variable	Statistic or Category	Placebo (n=76)	NBI 40 mg (n=72)	NBI 80 mg (n=79)	Total (N=227)
Age (years)	Mean	57	55	56	56
	SD	10.5	8.5	10.1	9.7
	Median	58	56	57	57
	Min, Max	30, 84	26, 74	32, 83	26, 84
Gender (n[%])	Male	42 (55)	42 (58)	39 (49)	123 (54)
	Female	34 (45)	30 (42)	40 (51)	104 (46)
Race* (n [%])	Black or African American	29 (38)	26 (36)	32 (41)	87 (38)
	Caucasian	43 (57)	41 (57)	44 (56)	128 (56)
	Disease Category (n[%])	Schizophrenia/ schizoaffective	50 (66)	48 (67)	52 (66)
	Mood Disorder	26 (34)	24 (33)	27 (34)	77 (34)

Baseline	Mean	9.9	9.7	10.4	10.0
AIMS	SD	4.3	4.1	3.6	4.0
Dyskinesia	Median	10	9	10	10
Total Score	Min, Max	0, 20	2, 20	3, 20	0, 20

(Source: Study Report p. 75-76; *omitted races with at most 1 patient per treatment group)

Study 1202

Dose Titration Summary

The majority of subjects in the Valbenazine and placebo groups had their dose up titrated at Week 2 (84% and 91%, respectively) and Week 4 (77% and 82%, respectively). A total of 34 Valbenazine subjects had their dose up titrated at both Week 2 (25 to 50 mg) and Week 4 (50 to 75 mg) compared with 35 placebo subjects. A summary of study drug titration is provided in Table 8 below.

Table 8. Study 1202 Summary of Dosing Titration by Visit and Treatment (Safety Analysis Set)

Visit	Category	Placebo N=49	Valbenazine N=51
Week 2	Maintain dose (n[%])	4 (9)	8 (16)
	Up titrate dose (n [%])	41 (91)	42 (84)
	Down titrate dose (n[%])	0 (0)	0 (0)
	Total (n)	45	50
Week 4	Maintain dose (n[%])	7 (16)	10 (21)
	Up titrate dose (n [%])	37 (82)	37 (77)
	Down titrate dose (n[%])	1 (2)	1 (2)
	Total (n)	45	48

(Source: Study Report p. 79)

3.2.4 Results and Conclusions

Sponsor's Results

Study 1202 Primary Efficacy Endpoint

Per Protocol Population

The primary efficacy endpoint (difference between Valbenazine and placebo groups in change from BL in AIMS Dyskinesia Total Score from baseline to week 6) was statistically significant for the per-protocol set (p-value < 0.0001).

Table 9. Study 1202 Primary Endpoint Results (Per-Protocol)

Study Number	Treatment Group	Primary Efficacy Measure: AIMS Dyskinesia Total Score (PP)		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine (25-75 mg/day)* [n=32]	8.0 (3.3)	-3.4 (1.2)	-3.0 (-4.5, -1.6)
	Placebo [n=44]	7.9 (4.5)	-0.3 (1.1)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

* Doses statistically significantly superior to placebo.

[Results confirmed by reviewer.]

Reviewer's note: The results for the Per Protocol population are listed here because it is the pre-specified primary analysis set for efficacy. This reviewer notes that a per protocol analysis is in principle not acceptable as primary efficacy analysis, because it violates the balance achieved by randomization.

ITT Population

Statistical significance for the primary efficacy endpoint was also achieved in the 'ITT' set (p-value: 0.0005).

Table 10. Study 1202 Primary Endpoint Results (ITT)

Study Number	Treatment Group	Primary Efficacy Measure: AIMS Dyskinesia Total Score (ITT)		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine (25-75 mg/day)* [n=45]	8.0 (3.5)	-2.6 (1.2)	-2.4 (-3.7, -1.1)
	Placebo [n=44]	7.9 (4.5)	-0.2 (1.1)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

* Doses statistically significantly superior to placebo.

[Results confirmed by reviewer.]

Reviewer's note: The 'ITT' set by the sponsor definition is equivalent to a completer set. Implications are explored in the reviewer's results section.

Secondary Efficacy Endpoint

The secondary efficacy endpoint Clinical Global Impression of Change-Tardive Dyskinesia (CGI-TD) achieved nominal statistical significance at week 6 (Table 12). CGI-TD was not pre-specified in a multiple testing procedure. Note that no patient in either group worsened over the course of the 6-week study (Table 11).

Table 11. Study 1202 CGI TD Assessment (Week 6) - ITT

Response Category	Placebo (n=44)	Valbenazine (n=45)
	n (%)	n (%)
<i>Very much improved</i>	2 (4.5)	6 (13.3)
<i>Much improved</i>	5 (11.4)	24 (53.3)
<i>Minimally improved</i>	24 (54.5)	12 (26.7)
<i>No change</i>	13 (29.5)	3 (6.7)
<i>Minimally worse</i>	0	0
<i>Much worse</i>	0	0
<i>Very much worse</i>	0	0

Table 12. Study 1202 Secondary Endpoint Results

Study Number	Treatment Group	Key Secondary Efficacy Measure: Clinical Global Impression of Change (Week 6)	
		LS Mean (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine (25-75 mg/day)	2.2 (0.3)	-0.8 (-1.2, -0.5)
	Placebo	3.1 (0.3)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean.

[Results confirmed by reviewer.]

Secondary results based on independent site raters

Note that the primary efficacy results for the AIMS Total Score are based on central raters scoring videotaped assessments. Prior to protocol amendment 2 the assessment and scoring was

performed by independent site raters. For completeness the independent site rater results are provided in Table 13. Note that they are not statistically significant.

Table 13. Study 1202 Secondary Endpoint Results (on site raters and ANCOVA)

Study Number	Treatment Group	Secondary Efficacy Measure: AIMS Dyskinesia Total Score (on-site raters) - ITT		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine (25-75 mg/day)	16.1 (4.7)	-4.3 (1.7)	-1.8 (-3.8, 0.1)
	Placebo	15.5 (5.2)	-2.4 (1.6)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Study Report p. 89; Results confirmed by reviewer.)

Supplementary Analysis (MMRM) on AIMS scores generated by site raters

The original protocol specified a repeated measures model using on site rater scores. The results of such analysis are provided in Table 14.

Table 14. Study 1202 Secondary Endpoint Results (on site raters and MMRM)

Study Number	Treatment Group	Secondary Efficacy Measure: AIMS Dyskinesia Total Score (on-site raters) - ITT		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine (25-75 mg/day)*	16.1 (4.7)	-5.1 (1.3)	-1.9 (-3.7, -0.1)
	Placebo	15.5 (5.2)	-3.2 (1.3)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

* Doses statistically significantly superior to placebo.

(Source: Study Report p. 90 and Sponsor Table 14.15.7; Results confirmed by reviewer.)

This approach would have resulted in statistically significant results at week 6 for the ITT population. The differences at weeks 2 and week 4 would not have represented statistical significant separations between the treatment arms.

Study 1304

Primary Efficacy Endpoint – AIMS

The primary efficacy endpoint Change from Baseline to Week 6 in AIMS Dyskinesia Total Score achieved statistical significance at $\alpha=0.05$ for the 80 mg Valbenazine dose ($p < 0.0001$). Table 15 presents the LS mean point estimate of the change and of the placebo-subtracted difference, with the patients treated with Valbenazine 80 mg on average improving by about 3 points more than the placebo treated patients. The 40 mg dose achieved nominal statistical significance ($p=0.0021$), but according to the sponsor specified multiple comparison Type I error control procedure testing could not proceed beyond the not statistically significant secondary endpoint (CGI-TD) for the 80 mg dose (Table 17).

Table 15. Study 1304 Primary Endpoint Results

Study Number	Treatment Group (n)	Primary Efficacy Measure: AIMS Dyskinesia Total Score (ITT)		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day) (70)	9.8 (4.1)	-1.9 (0.4)	-1.8 (-3.0, -0.7)
	Valbenazine (80 mg/day)* (80)	10.4 (3.6)	-3.2 (0.4)	-3.1 (-4.2, -2.0)
	Placebo (76)	9.9 (4.3)	-0.1 (0.4)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

* Doses statistically significantly superior to placebo.

(Results confirmed by reviewer.)

Key Secondary Efficacy Endpoint – CGI-TD

The key secondary endpoint CGI-TD LS mean comparison at week 6 did not result in statistically significant separation between either Valbenazine dose and placebo.

Table 16. Study 1304 Key Secondary Endpoint Results

Study Number	Treatment Group	Key Secondary Efficacy Measure: Clinical Global Impression of Change (Week 6)	
		LS Mean (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day)	2.9 (0.1)	-0.3 (-0.5, 0.0)
	Valbenazine (80 mg/day)	2.9 (0.1)	-0.3 (-0.5, 0.0)
	Placebo	3.2 (0.1)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean.

(Results confirmed by reviewer.)

Table 17. Study 1304 Summary Table of Efficacy Results given Multiplicity Control Procedure

Testing Procedure Sequence	P-value	Conclusion
Week 6 AIMS mean change from BL: Val 80 mg vs. placebo	<0.0001	Stat. significant.
Week 6 CGI-TD mean score: Val 80 mg vs. placebo	0.0560	Not stat. significant. Stop testing.
Week 6 AIMS mean change from BL: Val 40 mg vs. placebo	0.0021	Not stat. significant.
Week 6 CGI-TD mean score: Val 40 mg vs. placebo	0.0742	Not stat. significant.

(Source: Study Report p. 91)

Based on the pre-specified testing procedure, a statistically significant result has been observed for the primary endpoint for 80 mg dose. The Week 6 CGI-TD mean score for the 80 mg Valbenazine dose was not statistically different from placebo. Therefore all subsequent tests are declared not statistically significant.

Sensitivity Analysis Results – AIMS

The *tipping point* sensitivity analysis indicated that the AIMS dyskinesia total score mean difference for the 40 mg and 80 mg groups vs. placebo comparison at Week 6 remained significant (at nominal significance level of 0.05) up until the score for the Valbenazine treatment group subjects with missing Week 6 data were worsened by more than 60% and 230% respectively.

The *jump to reference* sensitivity analysis indicated statistically significant improvement (at nominal significance level of 0.05) in the AIMS dyskinesia total score in the 40 mg ($p=0.006$; 95% CI: -2.8, -0.5) and 80 mg ($p<0.0001$; 95% CI: -3.9, -1.6) groups vs. placebo at Week 6.

Open label extension of 1304

There were 202 ITT patients who completed the 6 week double-blind phase of the trial (Placebo: 69, NBI 40 mg: 63, and NBI 80 mg: 70). Of those 198 entered the open-label extension (with the placebo group subjects being re-randomized to 40 or 80 mg). A total of 124 (62.6%) subjects completed the extension period (week 48). The 74 (37%) discontinuations are mainly due to adverse events (15.7%), withdrawal of consent (8.6%), and being lost to follow-up (7.1%) (Table 18). AIMS assessments by central raters were performed at weeks 8, 16, 32, 48, and 52 (off drug) (Table 19, Figure 1).

One should be cautious when attempting to draw conclusions from the observed mean change from baseline AIMS scores in the extension phase given the substantial rate of discontinuations and the lack of a control group.

Table 18. Study 1304 Open-label Extension: Subject Enrollment and Disposition

	NBI 40 mg n (%)	NBI 80 mg n (%)	Total n (%)
Entered	97	101	198
Discontinued	36 (37.1)	38 (37.6)	74 (37.4)
Completed	61 (62.9)	63 (62.4)	124 (62.6)
Discontinuation Reason			
Adverse Event	14 (14.4)	17 (16.8)	31 (15.7)
Protocol Deviation	0 (0)	1 (1.0)	1 (0.5)
Non-Compliance	3 (3.1)	3 (3.0)	6 (3.0)
Withdrawal of Consent	9 (9.3)	8 (7.9)	17 (8.6)
Death	0 (0)	1 (1.0)	1 (0.5)
Lost to Follow-Up	8 (8.2)	6 (5.9)	14 (7.1)
Sponsor/Investigator Decision	2 (2.1)	2 (2.0)	4 (2.0)

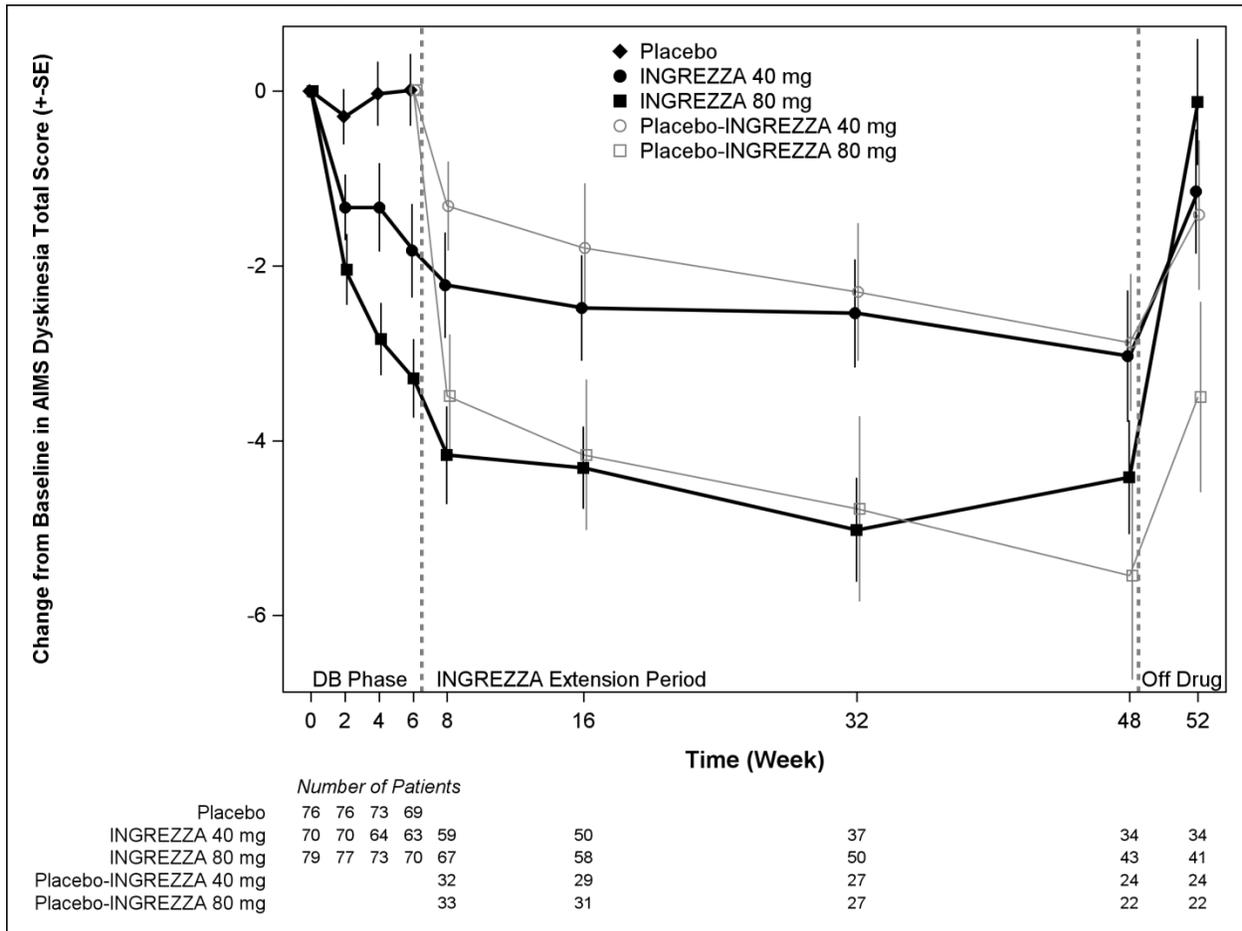
(Source: 120-Day Safety Update p. 1575)

Table 19. Study 1304 Observed Mean AIMS Total Score Change from Baseline During the Extension and Posttreatment Periods (ITT)

Visit	Statistic	NBI 40 mg (n=97)	NBI 80 mg (n=101)	N (n=198)
Week 8	n	94	97	191
	Mean (SEM)	-1.9 (0.4)	-4.0 (0.4)	
Week 16	n	81	87	168
	Mean (SEM)	-2.2 (0.5)	-4.3 (0.4)	
Week 32	n	66	75	141
	Mean (SEM)	-2.5 (0.5)	-4.9 (0.5)	
Week 48	n	60	63	123
	Mean (SEM)	-3.0 (0.5)	-4.8 (0.6)	
Week 52	n	60	61	121
	Mean (SEM)	-1.4 (0.5)	-1.2 (0.6)	
Follow-up	Mean (SEM)	-1.4 (0.5)	-1.2 (0.6)	

(Source: 120-Day Safety Update p. 33, 1576-1579)

Figure 1. Study 1304 AIMS Mean Score Change from Baseline During the Extension and Posttreatment Periods by Treatment (Arithmetic Mean \pm SEM, ITT)



(Source: Reviewer; compare to 120-Day Safety Update p. 33)

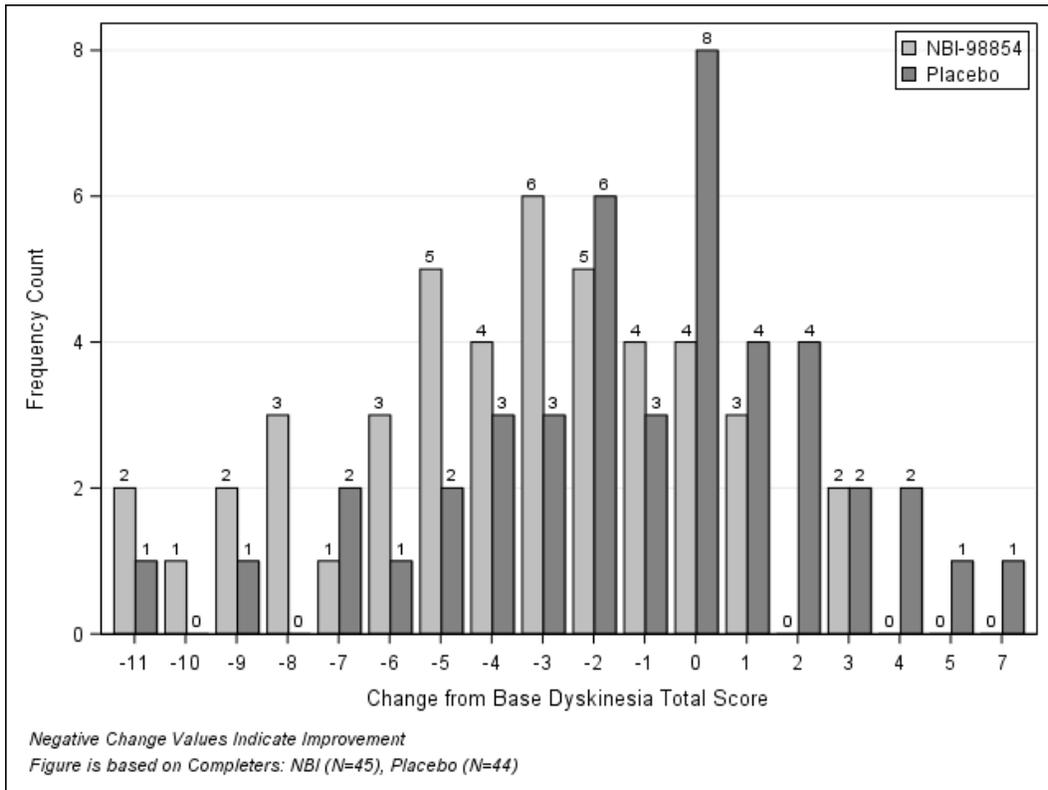
Reviewer's Results

This reviewer obtained the same results as presented by the sponsor for the primary and key secondary endpoints of Studies 1202 and 1304.

Study 1202

Figure 2 below displays the frequencies of the score changes observed for the AIMS total score by treatment group with negative score changes indicating improvement over the 6-week treatment period. The distributions overlap to some degree, but it is also apparent that the frequencies for Valbenazine treated patients are larger compared to placebo in the range indicating improvement.

Figure 2. Study 1202 Frequency of Subjects with Specified Magnitude in AIMS CFB at Week 6



(Source: Reviewer)

Review Issues

Study 1202

A host of review issues are due to the fact that Study 1202 is a Phase 2 (non-pivotal) study. As a consequence the clinical team was less worried about changes to the protocol and protocols and amendments were not reviewed by the Biometrics Division. A SAP appears to have never been submitted for review to the Division.

Review Issues for Study 1202

- 1) The Per Protocol Set (PP) is specified as the primary analysis set. The PP set excludes randomized subjects based on deviations from the protocol, which possibly affects the balance between treatment groups achieved by randomization. It is therefore in principle not acceptable as primary efficacy analysis set. The sponsor presents the efficacy results also for the 'ITT' set. The results for the PP and ITT sets (Table 9, Table 10) lead to the same conclusions favoring Valbenazine over placebo.
- 2) However, the original definition of the ITT set (i.e., baseline and at least one post-baseline AIMS score) is impacted by the switch from a repeated measures to an ANCOVA (change from BL to week 6) model with protocol amendment 2 and the decision to conduct central ratings only at baseline and week 6, which turns the 'ITT' set in a completer set. There are 12 subjects who discontinued during the DB phase and are excluded from the 'ITT' set.
- 3) No multiple testing control procedure (e.g., order in fixed testing sequence) was specified for the secondary endpoints CGI-TD and AIMS (independent site raters). As a consequence only the primary endpoint can be considered for labeling.
- 4) The change from independent site raters to central raters for the AIMS scoring occurred as study 1202 was already underway. Since the change occurred prior to unblinding of treatment assignment this reviewer is less concerned.
- 5) Titration design
Patients are up titrated from 25 mg, to 50 mg, to 75 mg given therapeutic response and tolerability/safety. At week 6 the mean dose is 64.4 mg/day. The number of patients on 25 mg, 50 mg, and 75 mg at week 6 are respectively 5, 9, and 31. The design makes it somewhat difficult to connect the results with the doses of 40 mg and 80 mg tested in the Phase 3 trial. Given that 69% of patients were titrated to 75 mg by week 6 and the favorable treatment effect for this subset of patients (Table 20 presents exploratory efficacy results by titration dose) Study 1202 appears supportive for the 80 mg dose.

Table 20. Study 1202 Primary Endpoint Results by Week 6 Dose

Study Number	Treatment Group (n)	Primary Efficacy Measure: AIMS Dyskinesia Total Score (ITT)		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1202	Valbenazine 25 mg/day (5)	7.8 (5.3)	-4.3 (1.7)	-4.1 (-7.0, -1.2)
	Valbenazine 50 mg/day (9)	9.1 (4.3)	-3.9 (1.5)	-3.7 (-5.9, -1.4)
	Valbenazine 75 mg/day (31)	7.7 (3.0)	-2.0 (1.2)	-1.8 (-3.2, -0.3)
	Placebo (44)	7.9 (4.5)	-0.2 (1.1)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Reviewer)

Caution needs to be exercised in interpreting the results shown in Table 20 especially for the 25 and 50 mg doses due to titration design (not fixed-dose design) and the small samples, since more difficult to treat patients get pushed to higher doses. Note that patients were not randomized to these four treatment groups.

The placebo-subtracted difference in AIMS Total Score Change from BL for 75 mg/day dose is -1.8 (somewhat smaller than the estimate of -2.4 for the combined Valbenazine doses) but fairly close.

A detailed evaluation of review issue 2): ‘ITT’ Set being Completer Set

There were 12 patients (Valbenazine: 5, Placebo: 7) who were randomized, but discontinued during the double-blind period (do not have a central rater AIMS score at week 6) and as consequence are not included in the ‘ITT’ set. Table 3 lists the discontinuation reasons (mostly noncompliance, withdrawal of consent, and adverse event). The numbers by treatment group are too small to determine any differences in discontinuation reasons between groups.

Figures A8 and A9 in the appendix display the AIMS profiles of those discontinued patients (AIMS scoring by independent site raters, since central raters only scored BL and Week 6). There appears to be an improvement on average in the AIMS total score for the discontinued

Valbenazine patients. Only one placebo patient who discontinued has an AIMS total score at week 2 or later (i.e., many discontinuations occurred early). There are three placebo patients with early termination visits that are not mapped to a visit week. Considering the available data it does not appear that the discontinued placebo subjects worsened on their AIMS scores prior to discontinuation. It is difficult to assess the impact of those exclusions from the ITT set because the AIMS data available (prior discontinuation) are from the independent site raters and the primary efficacy analysis is based on data generated by the central raters.

It is somewhat re-assuring though that the difference between the randomized set (N=102) and the 'ITT' set (N=89) is not that large and the discontinuations occurred roughly to the same degree in the Valbenazine and Placebo arms. Also patient profiles (as far data is available) do not seem to indicate large differences in the general efficacy trajectories (AIMS) for patients on Valbenazine or Placebo.

Exploratory Analysis ('Tipping Point')

To assess the potential impact of excluding randomized patients from the primary analysis set this reviewer performed a simple tipping point analysis by imputing incrementally worse scores for the excluded Valbenazine patients while imputing either no change or minimal improved scores for the excluded placebo subjects. The primary analysis was re-run with the excluded patients (scores imputed) added to the 'ITT' set. The discrepancy between the randomized set (N=102) and the 'ITT' set (N=89) is 13 (Valbenazine [n=6], Placebo [n=7]). Two slightly different scenarios were explored:

- 1) Assume placebo subjects maintain baseline AIMS score (CFB = 0)
- 2) Assume placebo patients improve slightly (CFB = -1)

A baseline score of 8 was imputed for the excluded patients from both treatment groups (the observed baseline mean for both treatment groups [ITT; central raters]). The results of this exploratory analysis are provided in Table 21.

Table 21. Study 1202 Sensitivity Analysis Including ‘Randomized but Excluded’ Patients

Assumed AIMS CFB		Placebo-subtracted Difference (95% CI)
Placebo (n=7)	Valbenazine (n=6)	
0	0	-2.3 (-3.5, -1.2)
0	1	-2.3 (-3.4, -1.1)
0	2	-2.2 (-3.4, -1.0)
0	5	-1.9 (-3.2, -0.6)
0	9	-1.5 (-3.0, -0.1)
0	10	-1.4 (-2.9, 0.1)
-1	0	-2.2 (-3.4, -1.0)
-1	8	-1.5 (-2.9, -0.1)
-1	9	-1.4 (-2.8, 0.1)

(Source: Reviewer)

To contextualize above results we need to compare the tipping points (CFB of 9 or 10) with the observed AIMS change scores of the Valbenazine group. Figure 2 displays the frequency distribution of AIMS CFB for the completers. The observed ranges (min, max) and means are for Placebo: (-11, 7) with mean CFB of -1.1 and for Valbenazine: (-11, 3) with mean CFB of -3.6.

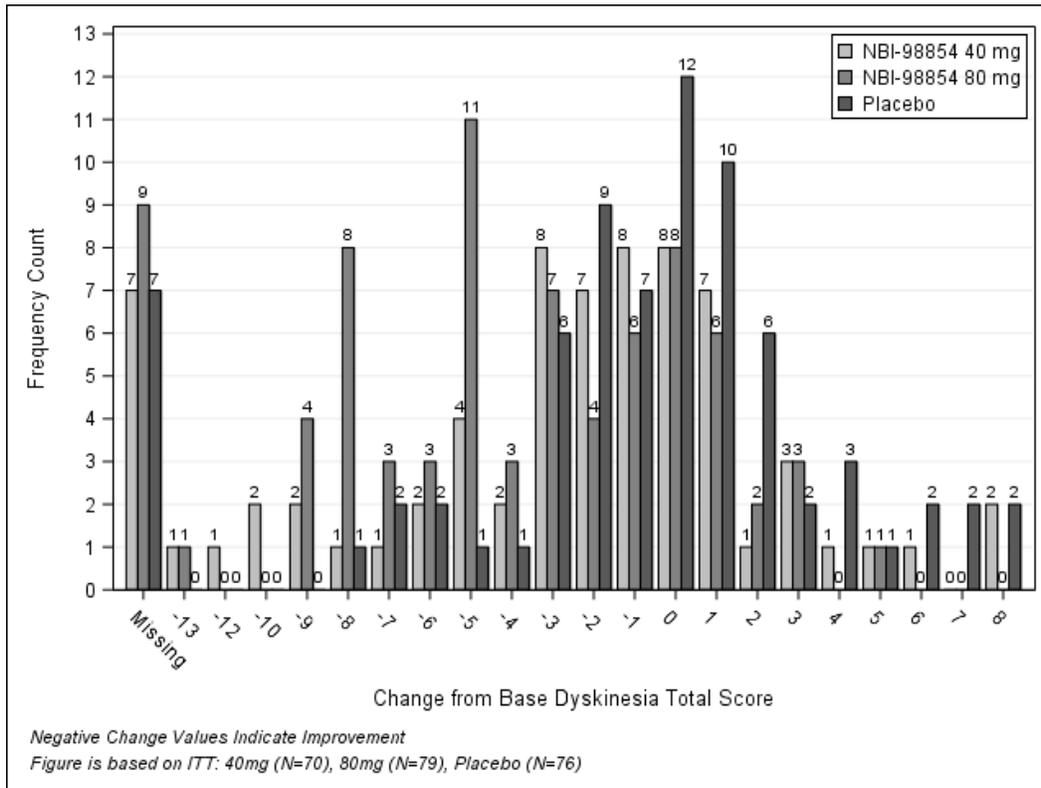
Conclusions

The six excluded Valbenazine patients would have to worsen on average by 9 or 10 points in the AIMS total score to turn the primary efficacy outcome to no longer statistically significant. Observing such magnitude of worsening seems rather unlikely given the highest observed worsening of three points for Valbenazine completers. A caveat of this simple approach is the omission of variability in the scores for excluded patients (i.e., imputation of the same score for each patient and not a realization from a distribution of scores.)

Study 1304

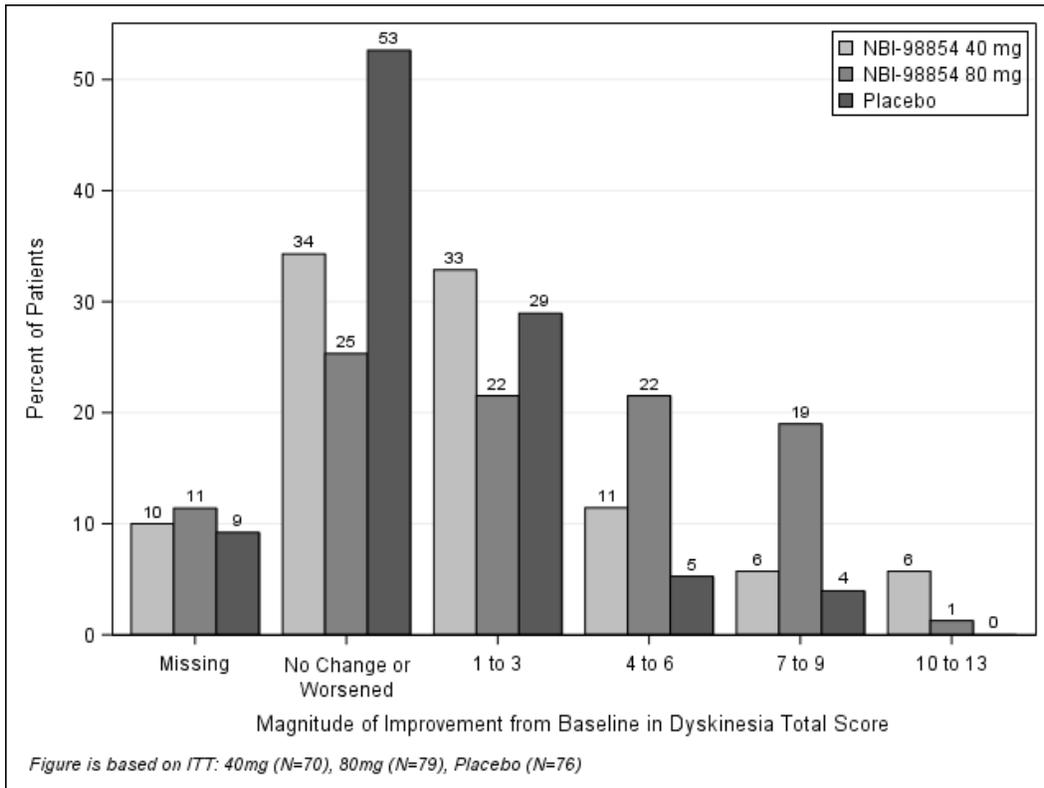
Figure 3 and Figure 4 below provide a visualization of the primary efficacy endpoint – change in AIMS Total Score. Patients on Valbenazine 80 mg more often realize a substantial improvement on the AIMS as compared to patients randomized to Valbenazine 40 mg or placebo.

Figure 3. Study 1304 Frequency of Subjects with Specified Magnitude in AIMS CFB at Week 6 (ITT)



(Source: Reviewer)

Figure 4: Study 1304 Percent of Patients with Specified Magnitude of AIMS Total Score Improvement at Week 6 (ITT)



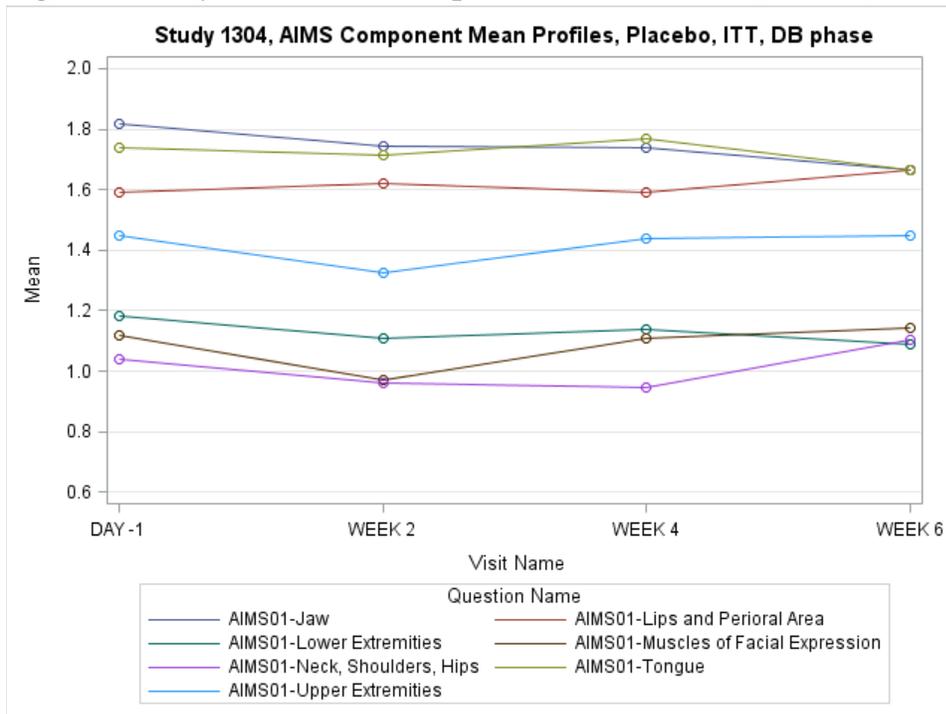
(Source: Reviewer)

For a figure displaying cumulative percentages of improvement (e.g., one point or better, two points or better ...) see figure A10 in the appendix.

Study 1304 AIMS Components

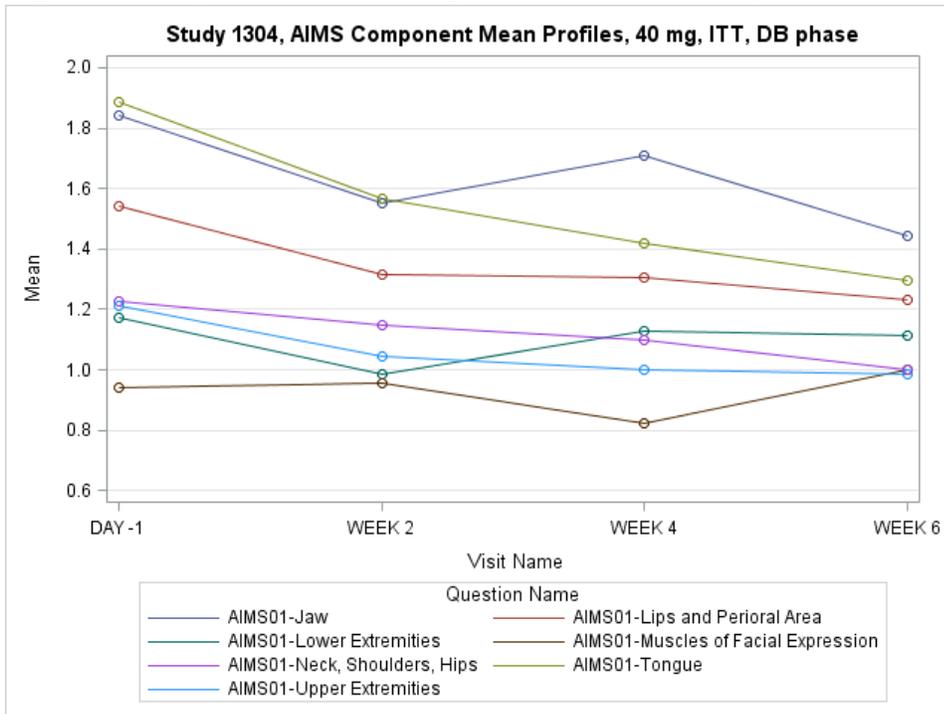
The AIMS Total Score is the sum of seven components (body regions). It is of interest to ascertain whether Valbenzine affects certain components more than others and whether the overall effect (i.e., reduction in total score) is driven by one or two strongly affected body regions. Figure 5 through Figure 7 display the mean AIMS component scores by week and within treatment group. It does not appear to be the case that the effect is concentrated in one or two body regions, all components of the AIMS appear to be affected by Valbenzine treatment, with the 80 mg dose showing the clearest trend for improvement.

Figure 5. Study 1304 AIMS Component Mean Profiles (Placebo)



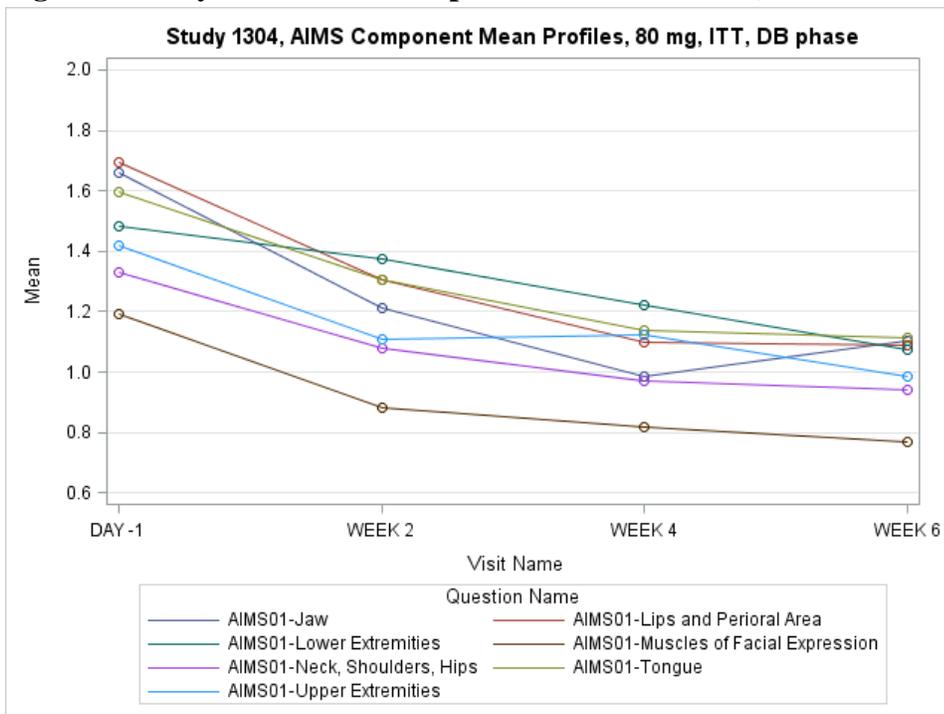
(Source: Reviewer)

Figure 6. Study 1304 AIMS Component Mean Profiles (Valbenazine 40 mg)



(Source: Reviewer)

Figure 7. Study 1304 AIMS Component Mean Profiles (Valbenazine 80 mg)



(Source: Reviewer)

3.3 Evaluation of Safety

The reader is referred to the clinical review for the evaluation of safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The purpose of the following subgroup analyses is to assess the consistency of treatment effects across subgroups. Randomization was not stratified by those subgroups, besides underlying disease category.

Study 1304

Gender

The proportion of males (53.8%) and females (46.2%) randomized into Study 1304 was similar (121 males versus 104 females). The results for the primary efficacy endpoint appear consistent across males and females (Table 22).

Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day)			
	Females (30)	10.2 (4.3)	-2.6 (0.7)	-1.8 (-3.6, -0.1)
	Males (40)	9.6 (4.0)	-1.3 (0.6)	-1.9 (-3.5, -0.4)
	Valbenazine (80 mg/day)			
	Females (40)	10.3 (3.7)	-3.3 (0.6)	-2.5 (-4.1, -0.9)
	Males (39)	10.4 (3.4)	-3.0 (0.6)	-3.6 (-5.1, -2.0)
	Placebo			
	Females (34)	10.4 (3.8)	-0.8 (0.6)	--
	Males (42)	9.5 (4.6)	0.6 (0.6)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Reviewer)

Race

The following is the break-down of the ITT set by race:

Black or African American (n=86 [38%]);

White (n=128 [57%]);

Other (n=6 [3%]);

American Indian or Alaska Native (n=2 [1%]);

Multiple (n=2 [1%]);

Native Hawaiian or other Pacific Islander (n=1 [0.5%]).

This reviewer limited the exploratory subgroup analysis by race to ‘Black or African American’ and ‘White’ race, due to the minimal representation of any other race in Study 1304. African Americans appear to see no effect of Valbenzazine when on the 40 mg dose. Due to the small sample size this finding could have occurred by chance.

Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenzazine (40 mg/day)			
	African American (25)	9.5 (4.8)	-1.0 (0.7)	-0.1 (-2.1, 1.8)
	White (41)	10.1 (3.7)	-2.4 (0.6)	-2.6 (-4.2, -1.1)
	Valbenzazine (80 mg/day)			
	African American (32)	9.6 (3.5)	-3.4 (0.7)	-2.5 (-4.3, -0.6)
	White (44)	11.0 (3.6)	-2.9 (0.5)	-3.2 (-4.7, -1.7)
	Placebo			
	African American(29)	9.7 (4.0)	-0.9 (0.7)	--
	White (43)	10.2 (4.3)	0.2 (0.5)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Reviewer)

Age

A total of 36 (16%) patients randomized to Study 1304 were 65 years or older, the other 189 (84%) were younger than 65. Eight of the 36 were randomized to Valbenazine 40 mg, 12 to Valbenazine 80 mg and 16 to Placebo.

Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day)			
	< 65 (62)	9.9 (4.2)	-1.9 (0.5)	-1.7 (-3.0, -0.5)
	≥ 65 (8)	9.3 (3.2)	-1.7 (1.3)	-2.1 (-5.1, 1.0)
	Valbenazine (80 mg/day)			
	< 65 (67)	10.3 (3.5)	-3.3 (0.4)	-3.2 (-4.4, -1.9)
	≥ 65 (12)	10.5 (4.1)	-2.1 (1.1)	-2.5 (-5.2, 0.2)
	Placebo			
	< 65 (60)	9.5 (4.1)	-0.2 (0.5)	--
	≥ 65 (16)	11.6 (4.7)	0.4 (0.9)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Reviewer)

The treatment effect appears similar for the two age groups. Any inferences for the older age bracket should be regarded cautiously in light of the small sample size.

Study 1202

Subgroup analyses were not performed for Study 1202 due to small overall sample size ('ITT' population of 89) and the dose titration design.

4.2 Other Special/Subgroup Populations

Exploratory subgroup analyses were also conducted based on metabolizer status, disease category, and antipsychotic medication use.

Genotype/Metabolizer Type

Table 25: Study 1304 Metabolizer Subgroup Frequencies

Genotype	Frequency (%)
CYP2D6 Extensive Metabolizer	125 (55.8)
CYP2D6 Extensive Metabolizer or Intermediate Metabolizer	4 (1.8)
CYP2D6 Extensive Metabolizer or Ultra Rapid Metabolizer	5 (2.2)
CYP2D6 Intermediate Metabolizer	70 (31.3)
CYP2D6 Poor Metabolizer	13 (5.8)
CYP2D6 Ultra Rapid Metabolizer	7 (3.1)

Table 26. Study 1304 Exploratory Subgroup Analysis by CYP2D6 Metabolizer Category

Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day)			
	Intermediate M* (31)	10.3 (4.6)	-1.3 (0.6)	-1.9 (-3.8, 0.1)
	Extensive M** (39)	9.4 (3.7)	-2.4 (0.6)	-2.0 (-3.5, -0.6)
	Valbenazine (80 mg/day)			
	Intermediate M (29)	10.2 (3.2)	-3.6 (0.7)	-4.2 (-6.2, -2.2)
	Extensive M (50)	10.5 (3.8)	-2.9 (0.5)	-2.6 (-4.0, -1.2)
	Placebo			
	Intermediate M (23)	10.0 (4.3)	0.6 (0.7)	--
	Extensive M (53)	9.9 (4.3)	-0.3 (0.5)	--

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline.

*Intermediate M includes patients classified as: Intermediate (n=70) and Poor (n=13) CYP2D6 Metabolizer.

** Extensive M includes patients classified as: Extensive (n=125), Extensive or Intermediate (n=4), Extensive or Ultra Rapid (n=5), and Ultra Rapid (n=7) CYP2D6 Metabolizer (n=7). (Source: Reviewer)

Patients on Valbenazine 80 mg categorized as intermediate metabolizers appear to realize a somewhat larger effect on their TD symptoms as compared to the extensive metabolizer group.

Disease Category

Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference ^a (95% CI)
Study 1304	Valbenazine (40 mg/day)			
	Mood disorder* (24)	11.4 (3.5)	-1.9 (0.7)	-1.7 (-3.7, 0.3)
	Schizophrenia** (46)	9.0 (4.2)	-1.9 (0.5)	-1.9 (-3.4, -0.5)
	Valbenazine (80 mg/day)			
	Mood disorder (27)	10.9 (3.8)	-3.1 (0.7)	-2.9 (-4.8, -1.0)
	Schizophrenia (52)	10.1 (3.5)	-3.2 (0.5)	-3.2 (-4.7, -1.8)
Placebo				
Mood disorder (26)	11.2 (3.6)	-0.2 (0.7)	--	
Schizophrenia (50)	9.3 (4.5)	0.0 (0.5)	--	

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

^a Difference (drug minus placebo) in least-squares mean change from baseline. *Mood disorder with neuroleptic-induced TD.

**Schizophrenia or schizoaffective disorder with neuroleptic-induced TD.

(Source: Reviewer)

The effects of treatment appear similar across disease categories. Note that the categorization as “schizophrenia/schizoaffective” and “mood disorders” is somewhat questionable as far as clinically meaningfulness is concerned. Mood disorders include depression and bipolar disorder, and bipolar disorder is more similar to schizoaffective disorder than unipolar depression (Mike Davis, FDA clinical reviewer).

Antipsychotic Use

Table 28. Study 1304 Exploratory Subgroup Analysis by Antipsychotic Medication Use				
Study Number	Treatment Group (n)	Efficacy Measure: Change from Baseline in AIMS Dyskinesia Total Score at Week 6		
		Mean Baseline Score (SD)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference^a (95% CI)
Study 1304	Valbenazine (40 mg/day)			
	No (7)	12.6 (3.3)	-1.0 (1.4)	-1.3 (-4.6, 2.1)
	Yes (63)	9.5 (4.1)	-2.0 (0.5)	-1.9 (-3.1, -0.6)
	Valbenazine (80 mg/day)			
	No (16)	12.0 (4.4)	-3.7 (0.9)	-3.9 (-6.5, -1.4)
	Yes (63)	10.0 (3.2)	-3.0 (0.5)	-2.9 (-4.2, -1.6)
Placebo				
No (13)	9.8 (4.0)	0.2 (1.0)	--	
Yes (63)	10.0 (4.4)	-0.1 (0.5)	--	

SD: standard deviation; SE: standard error; LS Mean: least-squares mean; CI: confidence interval.

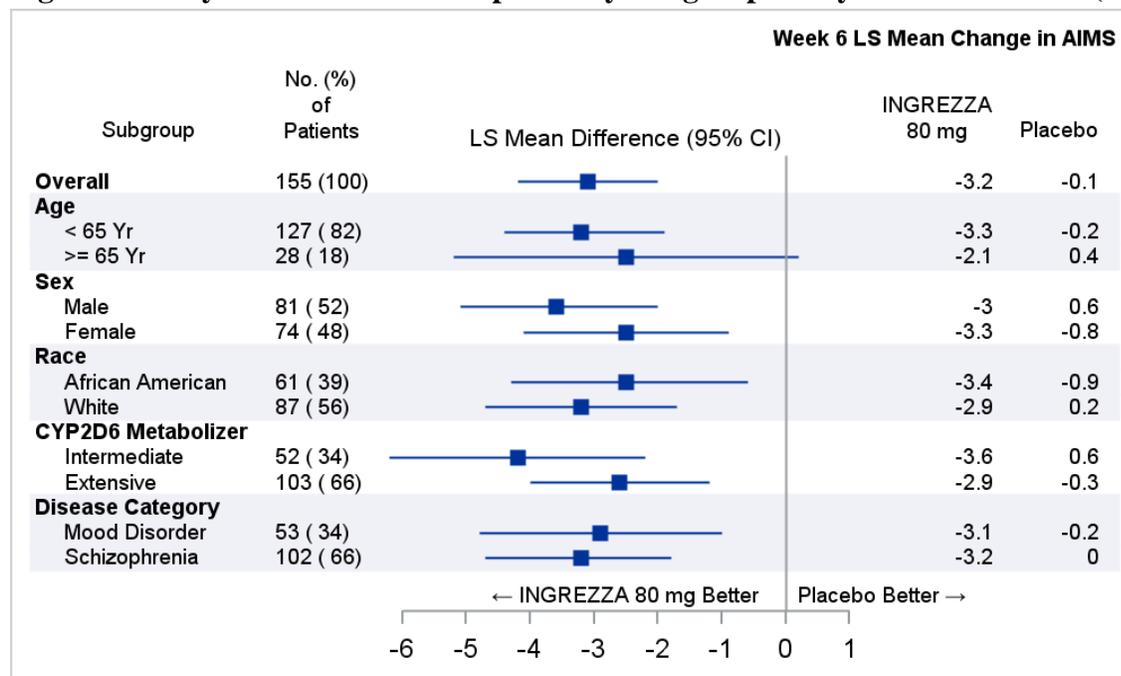
^a Difference (drug minus placebo) in least-squares mean change from baseline.

(Source: Reviewer)

Only a small subset of subjects was not on a stable antipsychotic upon entry into the study. The small sample size prevents any firm conclusions.

A summary of the exploratory subgroup analyses for the Valbenazine 80 mg group versus placebo is provided in Figure 8 below.

Figure 8. Study 1304 Forest Plot Exploratory Subgroup Analysis of AIMS CFB (80 mg)



(Source: Reviewer)

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

The basic issue in this NDA is the approach by the sponsor to put forward one *Phase 2 dose titration study* and one Phase 3 multiple dose study to support their claim.

There are essentially two problems arising from a study labeled as Phase 2 and designed as a dose titration study when submitted as the second adequate and well controlled trial under the NDA. The first is the difficulty to directly derive support for either of the two fixed doses tested in Phase 3 (40 mg or 80 mg) from the dose titration study.

The second is that by its nature of being labeled as Phase 2, the regulatory requirements communicated to the sponsor during the IND stage were lower. In particular, Study 1202 protocol amendment 2 (change to central raters and ANCOVA) did not receive the same level of scrutiny as a study labeled as pivotal would have received. Especially important for this statistical reviewer, protocol amendment 2 did never receive a Biometrics review. In summary, the strength of statistical evidence to support an efficacy indication should not be compromised by a trial that was initially considered exploratory unless the protocol and SAP were reviewed and agreed upon with the same rigor as applied to a pivotal trial.

Study 1202

With the majority of subjects being titrated to the 75 mg dose at the end of Study 1202 and this subset showing a treatment effect in an exploratory analysis this reviewer is satisfied as far as replication of the positive result of the 80 mg dose in the Phase 3 study is concerned.

Changes in the planned analysis from the original protocol (MMRM to ANCOVA) led to the 'ITT set' being a completer set. Fortunately in this therapeutic setting the number of discontinuations during the double-blind phase was modest. An exploration of the patient efficacy trajectories prior discontinuation and the results of a tipping point analysis lend some confidence to the results obtained from the 'ITT'/Completer set.

Study 1304

The sponsor specified multiple testing Type I error control procedure did not allow for the inferential testing of the primary efficacy endpoint for the 40 mg dose (nominal p-value of 0.002). The Division had advised the sponsor to change the proposed procedure, but the sponsor declined that proposal.

5.2 Collective Evidence

The Phase 3 Study 1304 provides strong statistical evidence for the 80 mg dose on the primary efficacy endpoint. The primary efficacy endpoint for the 40 mg dose cannot be declared statistically significant due to the failure of the 80 mg dose on the secondary efficacy endpoint, which is located above in the fixed testing sequence.

The Phase 2 Study 1202, despite numerous flaws, is solid enough to serve as replication of the positive findings for the 80 mg dose.

5.3 Conclusions and Recommendations

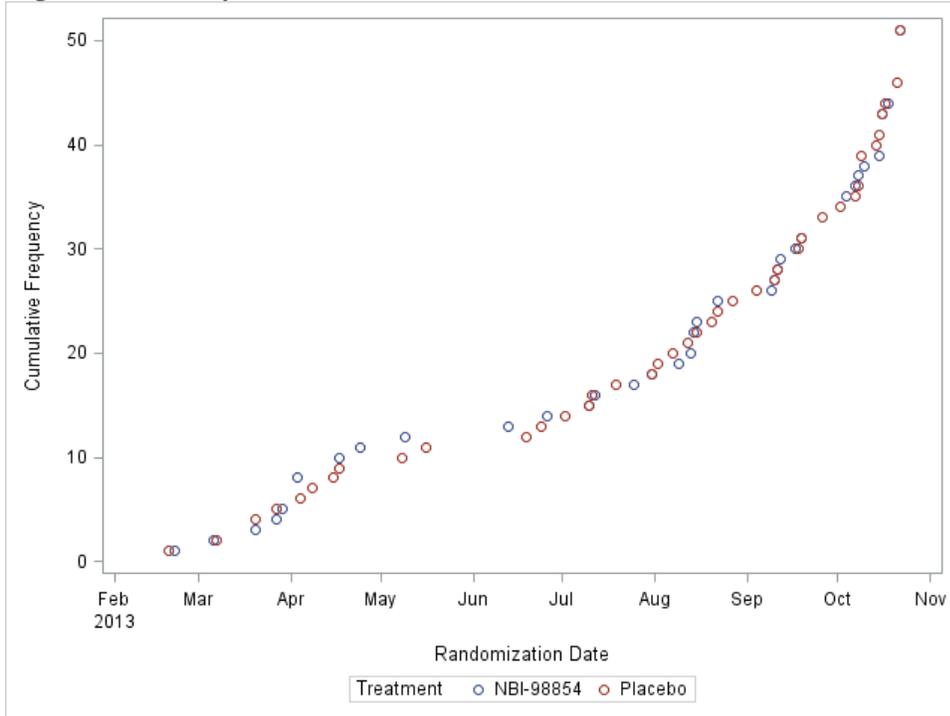
The statistical results provide adequate evidence to support the Valbenazine 80 mg dose for the acute treatment of Tardive Dyskinesia. The Valbenazine 40 mg dose appears to benefit some patients with TD; however it did not meet the strict evidentiary standard for efficacy (e.g., replication of finding).

5.4 Labeling Recommendations

(b) (4) need to be removed from (b) (4)

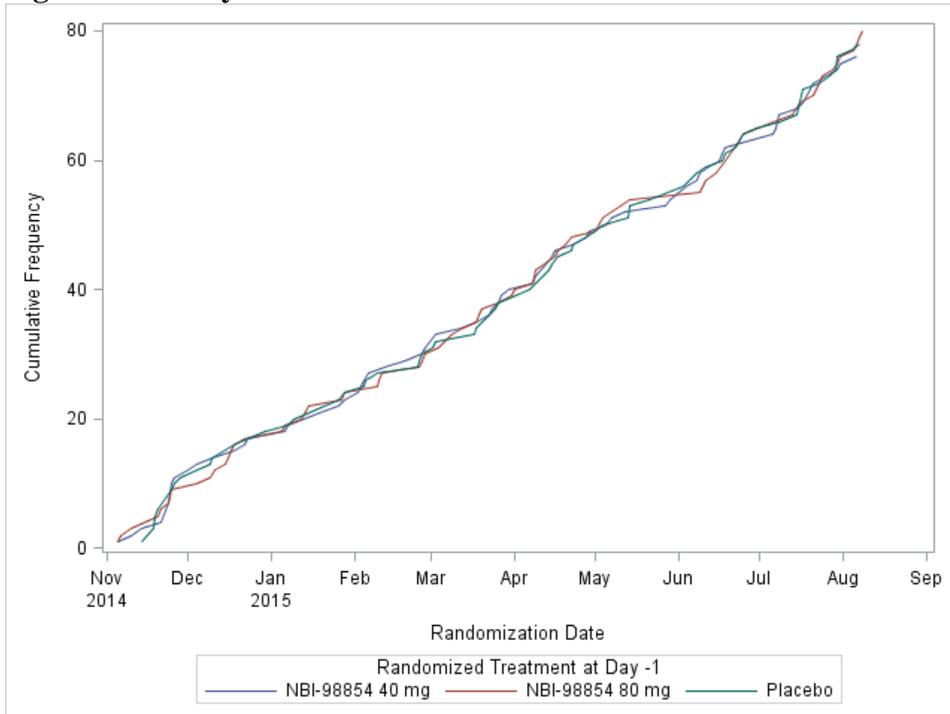
APPENDICES

Figure A1. Study 1202 Randomization



(Source: Reviewer)

Figure A2. Study 1304 Randomization



NDA 209,241 Site Selection Funnel Plots

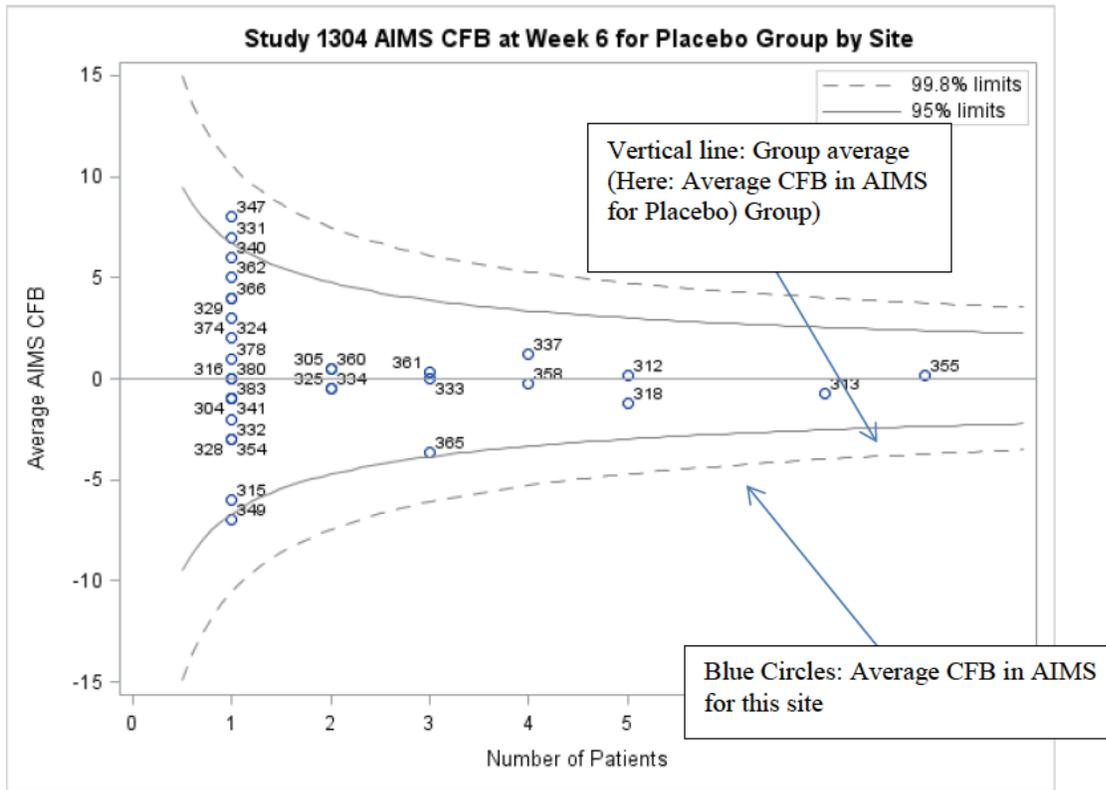
Study 1304

Table A1. Sites with highest enrollment (Study 1304*)

Siteid	Investigator	Patients enrolled	Audited by sponsor
312	Kenia Castro (Hialeah, FL)	12	Yes
313	Julio Castro-Gayol (Hialeah, FL)	12	Yes
318	Bernadette D'Souza	9	Yes
335	Eptesam A Khaled	7	
337	Mary Ann Knesewich	13	
355	Dolores Sanchez-Cazau (Hialeah, FL)	14	
358	Rajinder Shiwach	13	Yes
361	Cherian Verghese	10	Yes
374	Armen Krikor Goenjian	6	

*63 sites total

Figure A3



CFB = Change from Baseline, AIMS = Abnormal Involuntary Movement Scale

Figure A4

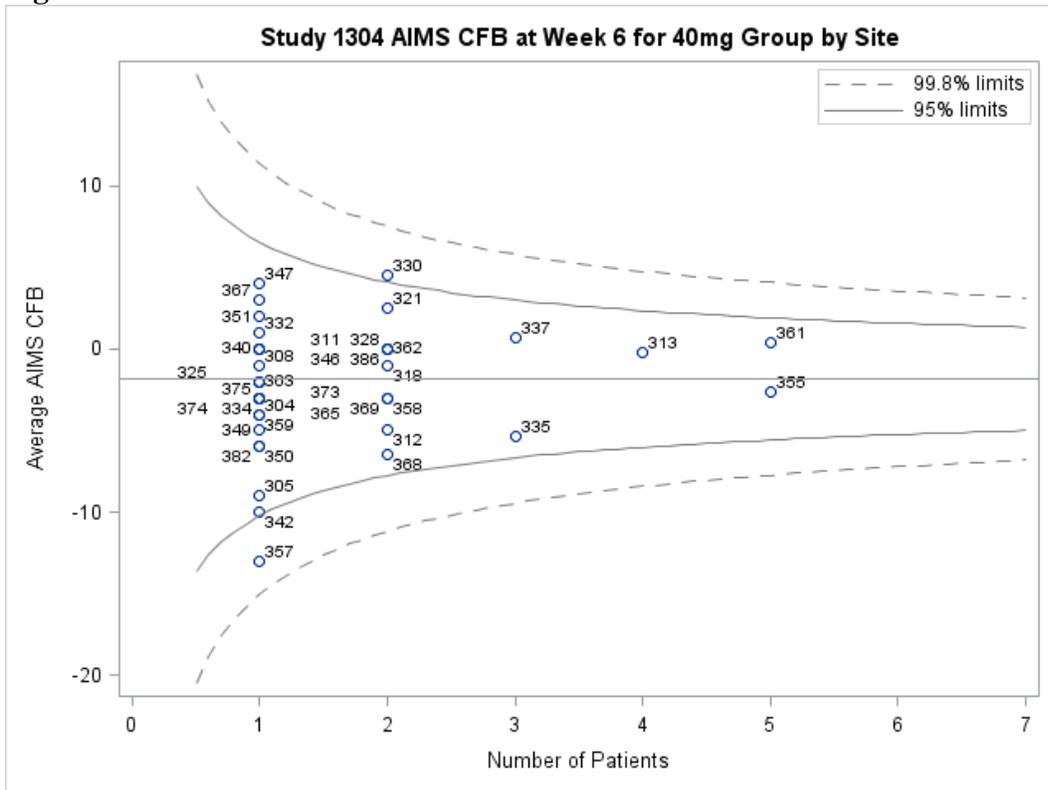
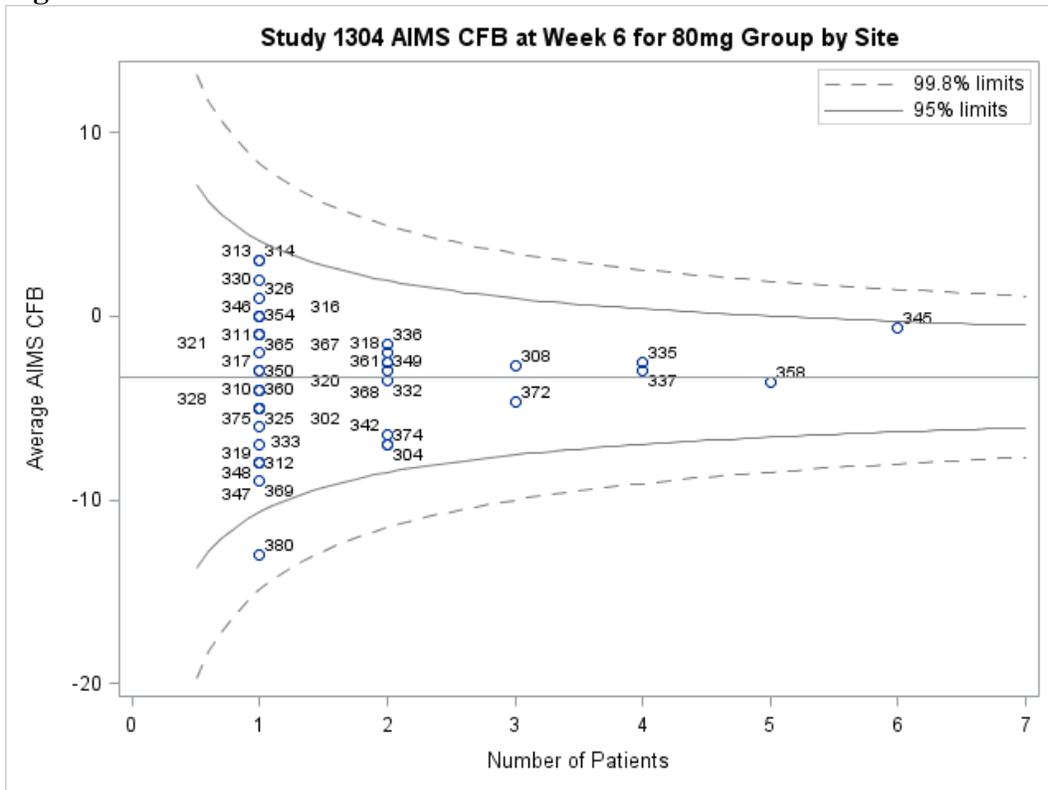


Figure A5



Study 1202

Table A2. Sites with highest enrollment (Study 1202*)

Siteid	Investigator	Patients enrolled	Audited by sponsor
204	Daniel F. Mantri	11	
215	Kenia Castro (Hialeah, FL)	17	
224	Julio Castro-Gayol (Hialeah, FL)	18	Yes

*29 sites total (the other 26 sites enrolled ≤ 5 subjects each)

Figure A6

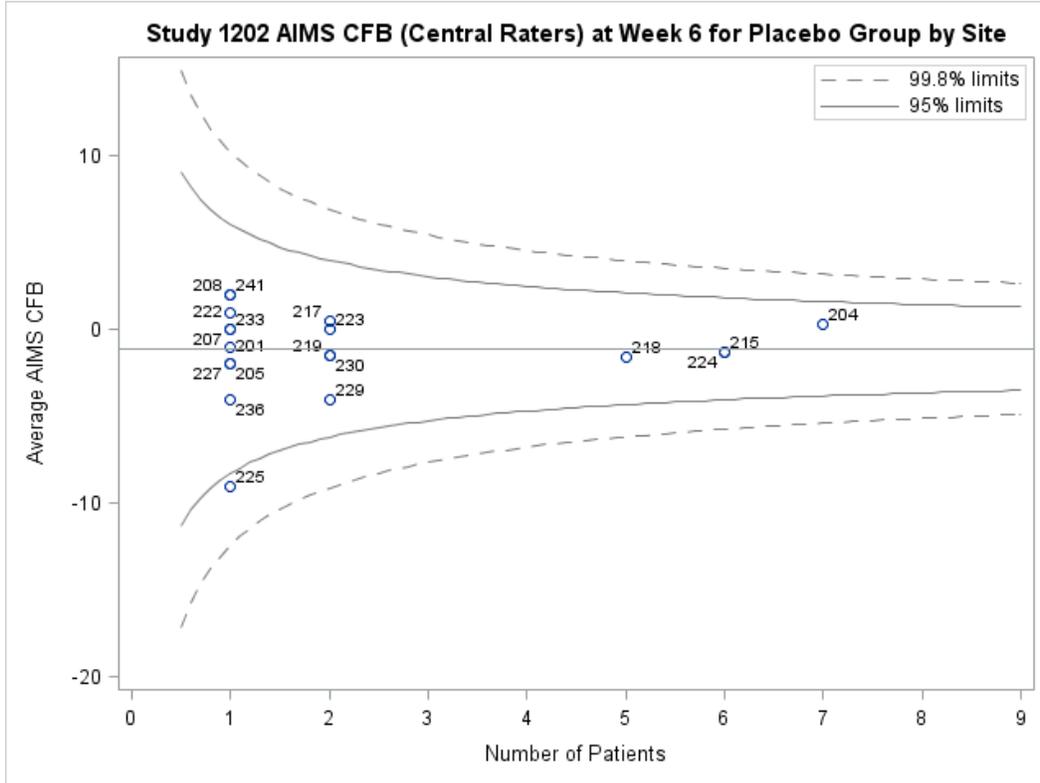
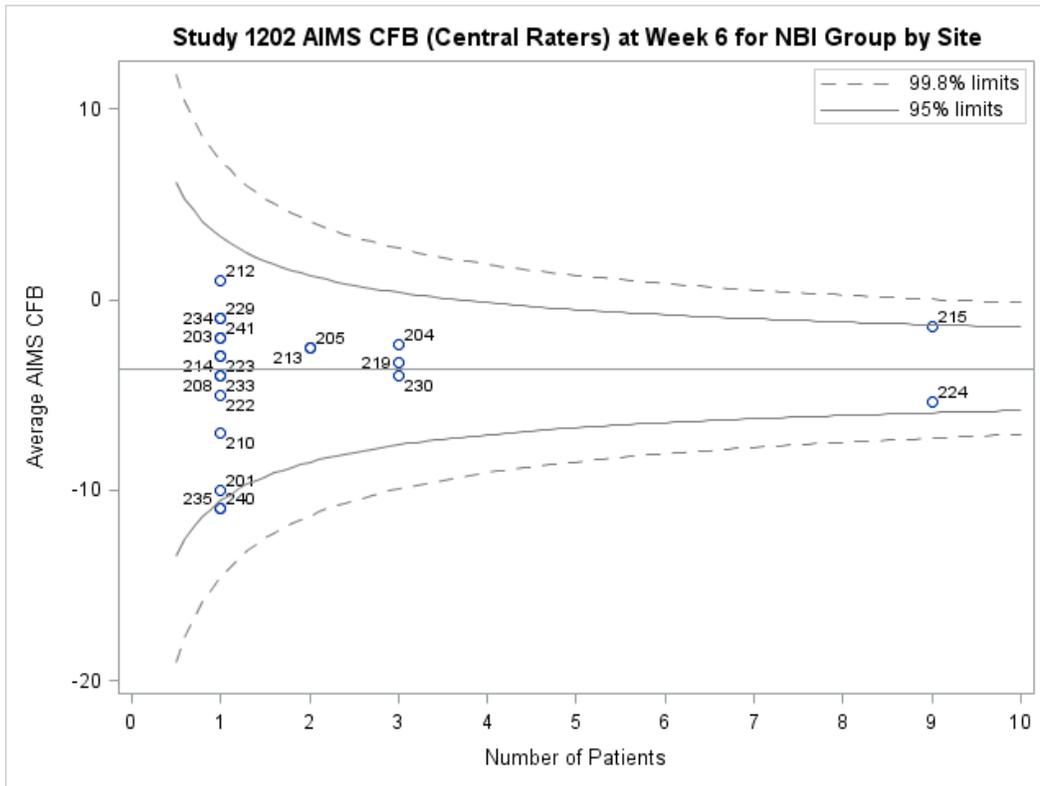


Figure A7



Conclusions:

From the funnel plots: There are no really extreme sites (i.e., sites whose average CFB falls outside of the 95% or even 99.8% confidence bounds of the mean) with more than two subjects forming the site average.

Figure A8. Study 1202 AIMS Total Score (site rating) Patient Profiles for Valbenazine Subjects who Discontinued Early from DB Phase

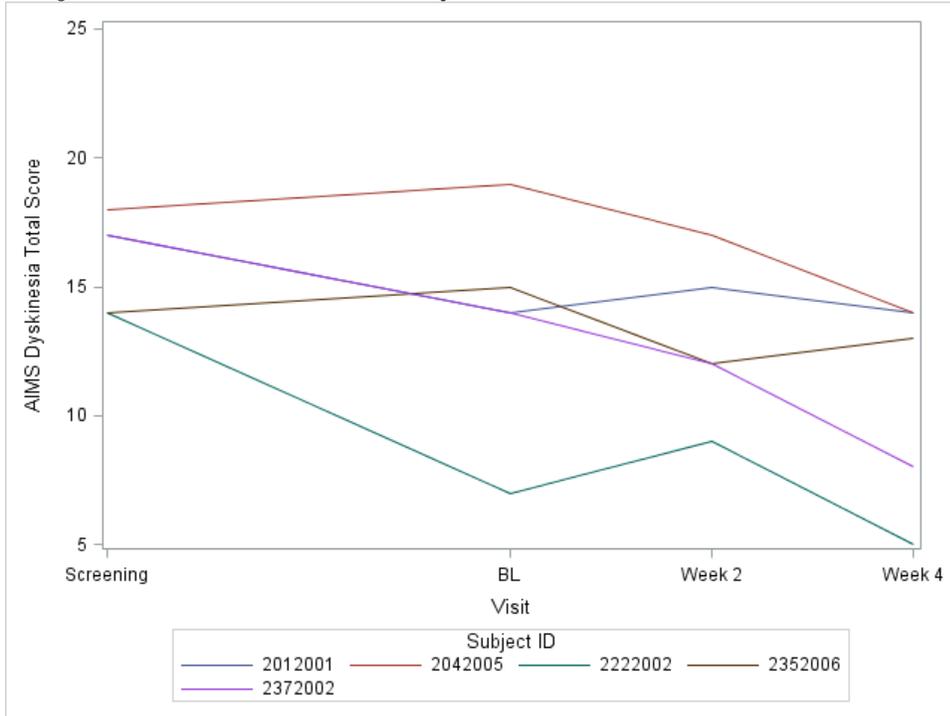
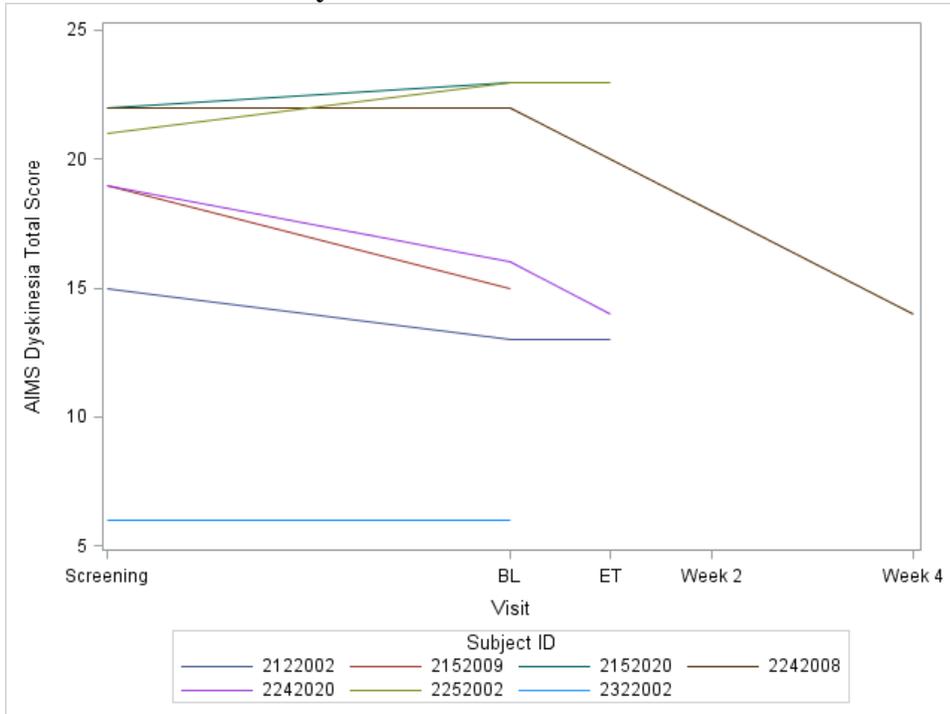


Figure A9. Study 1202 AIMS Total Score (site rating) Patient Profiles for Placebo Subjects who Discontinued Early from DB Phase



(ET = Early Termination)

Table A3. Study 1202 Mean AIMS Dyskinesia Total Scores (site rating) of Patients excluded from 'ITT'

Visit	Placebo	Valbenazine
	n Mean	n Mean
Screening	7 17.7	5 16.0
Baseline	7 16.9	5 13.8
Early Term*	3 16.7	
Week 2	1 18.0	5 13.0
Week 4/ET	1 14.0	5 10.8

(*Early Term = Early Termination. Note that the three early termination visits not necessarily occurred between the Baseline and prior to the scheduled Week 2 visits.)

Figure A10

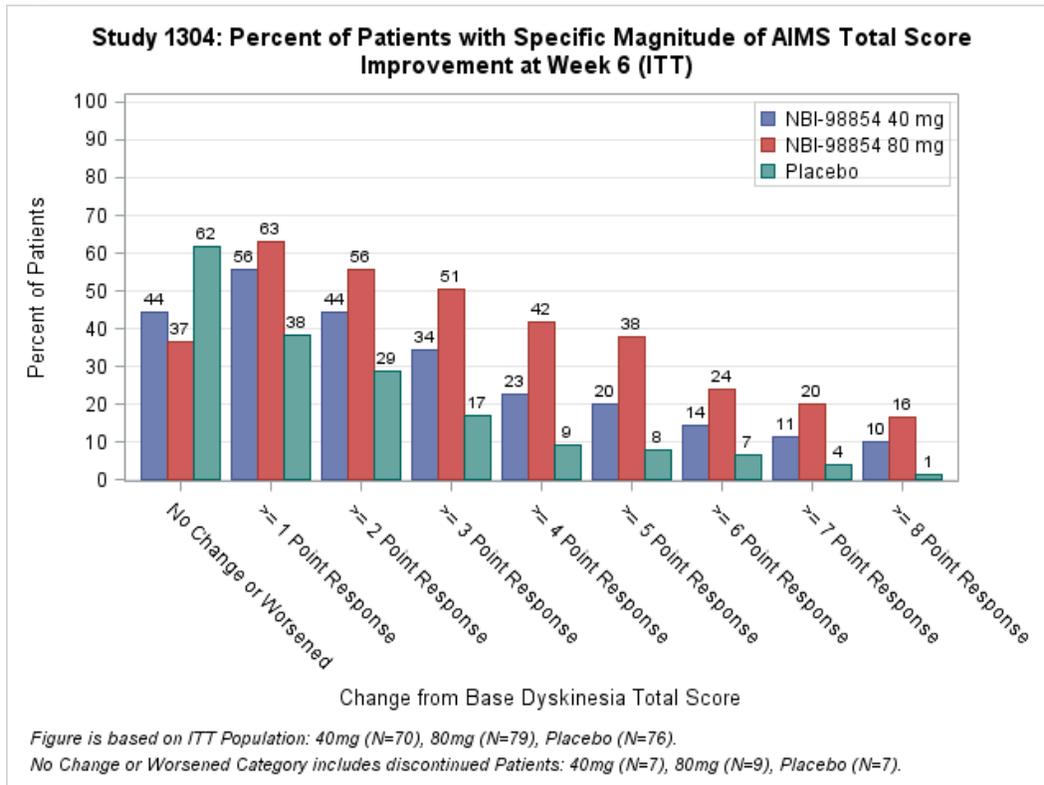


Figure A11. Study 1304 AIMS CFB by Gender

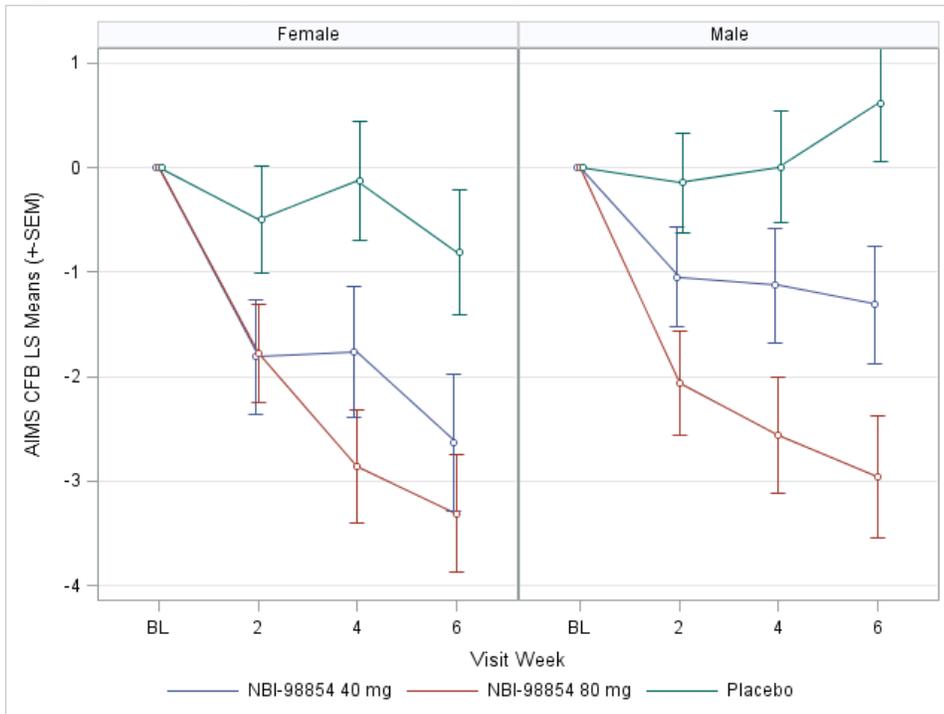
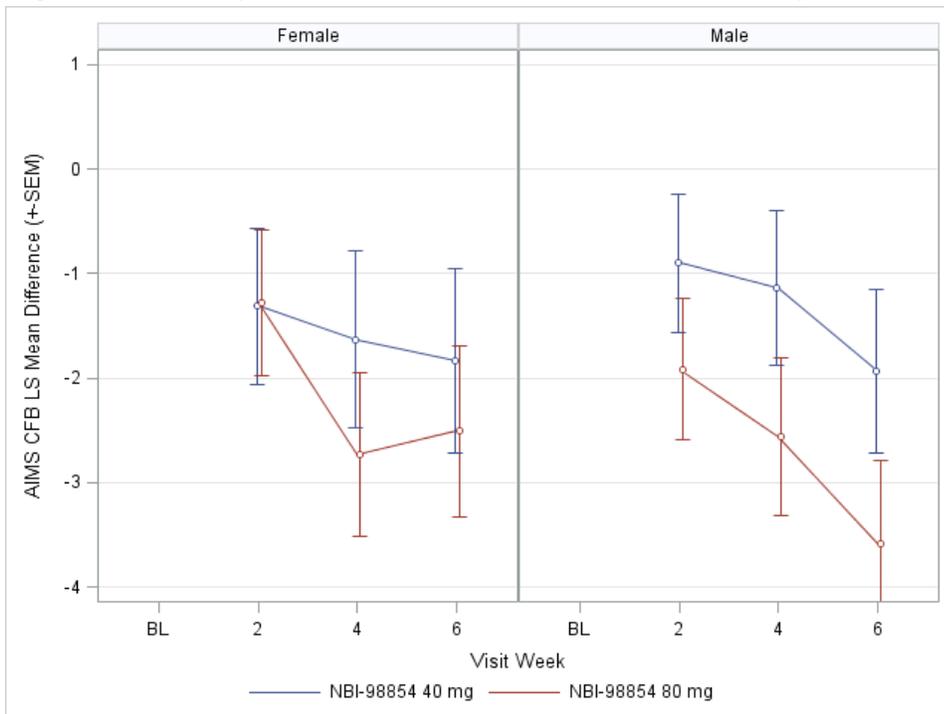


Figure A12. Study 1304 AIMS CFB LS Mean Difference by Gender



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

/s/

THOMAS BIRKNER
03/03/2017

PEILING YANG
03/04/2017

HSIEN MING J HUNG
03/06/2017