

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
NDA 21-894

STATISTICAL REVIEW(S)



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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 21, 894 (N_000)

Drug Name: Tetrabenazine

Indication(s): Huntington's Disease Chorea

Applicant: Prestwick Pharmaceuticals

Date(s): Submission Date: September 23, 2005

Review Priority: Priority

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Tristan Massie, Ph.D.

Concurring Reviewers: Kun Jin, Ph.D., Team Leader
Kooros Mahjoob, Ph.D., Deputy Director, Division of Biometrics 1

Medical Division: Neurologic Drugs (HFD-120)

Clinical Team: Carole Davis, M.D.
John Feeney, M.D., Team Leader
Russell Katz, M.D., Division Director

Project Manager: Teresa Wheelous

Keywords: Analysis of Covariance; Secondary variables; Standard of evidence;
Staggered Withdrawal Study; Study Conduct Error

Table of Contents

| | |
|---------------------------------------------------------------------------------------|-----------|
| LIST OF TABLES..... | 3 |
| LIST OF FIGURES..... | 4 |
| 1 EXECUTIVE SUMMARY | 5 |
| 1.1 CONCLUSIONS AND RECOMMENDATIONS | 5 |
| 1.2 BRIEF OVERVIEW OF CLINICAL STUDIES | 5 |
| 1.3 STATISTICAL ISSUES AND FINDINGS | 6 |
| 2 INTRODUCTION | 8 |
| 2.1 OVERVIEW..... | 8 |
| 2.2 DATA SOURCES | 8 |
| 3 STATISTICAL EVALUATION | 9 |
| 3.1 EVALUATION OF EFFICACY | 9 |
| 3.1.1 <i>Study TBZ 103,004</i> | 9 |
| 3.1.1.1 Study Design | 9 |
| 3.1.1.2 Efficacy Measures | 10 |
| 3.1.1.3 Statistical Methods and Sample size | 10 |
| 3.1.1.4 Disposition of Patients | 11 |
| 3.1.1.5 Patient Demographics | 12 |
| 3.1.1.6 Sponsor's Results..... | 13 |
| 3.1.1.7 Reviewer's Results..... | 16 |
| 3.1.1.7.1 Primary Analysis..... | 16 |
| 3.1.1.7.2 Assessment of the Impact of Missing Data | 16 |
| 3.1.1.7.3 Secondary Analyses | 17 |
| 3.1.1.7.4 Summary of Secondary Endpoints Results | 22 |
| 3.1.1.7.5 Group Differences on Change in Maximal Chorea within Individual Sites | 25 |
| 3.1.2 <i>Study TBZ 103,005</i> | 26 |
| 3.1.2.1 Study Design | 26 |
| 3.1.2.2 Efficacy Measures | 26 |
| 3.1.2.3 Statistical Methods | 27 |
| 3.1.2.4 Patient Disposition | 27 |
| 3.1.2.5 Patient Demographics | 27 |
| 3.1.2.6 Sponsor's Results..... | 28 |
| 3.1.2.7 Reviewer's Results..... | 30 |
| 3.2 EVALUATION OF SAFETY | 31 |
| 4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS | 32 |
| 4.1 GENDER, RACE AND AGE | 32 |
| 4.2 OTHER SPECIAL/SUBGROUP POPULATIONS | 34 |
| 5 SUMMARY AND CONCLUSIONS | 34 |
| 5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE | 34 |
| 5.2 CONCLUSIONS AND RECOMMENDATIONS | 36 |
| APPENDIX – INDIVIDUAL ITEMS OF THE FUNCTIONAL ASSESSMENT CHECKLIST | 37 |

LIST OF TABLES

| | |
|-----------------------------------------------------------------------------------------------------------------|----|
| Table 1 Study Characteristics | 5 |
| Table 2 TBZ103, 004: Baseline Demographic Characteristics and Baseline Efficacy Measures..... | 12 |
| Table 3 TBZ 103,004: Primary Efficacy Analysis Adjusted Mean Change (\pm S.E.M) in Total Chorea Score | 13 |
| Table 4 TBZ 103,004: Dose at Week 7 and Percent with 3 point decrease from baseline in Total Chorea score..... | 15 |
| Table 5 TBZ 103,004: Mean Total Chorea Scores at Baseline, Week 12 and Week 13 (1 week after withdrawal) ... | 16 |
| Table 6 TBZ 103,004: Sensitivity Analyses of Primary Endpoint | 17 |
| Table 7 TBZ 103,004: Adjusted Mean Change from Baseline in UHDRS Part II - Cognitive Assessment Items..... | 23 |
| Table 8 TBZ 103,004: Adjusted Mean Change (Average Week 9 and Week 12) from Baseline in Endpoints | 24 |
| Table 9 TBZ 103,005 Distribution of Stable Dose of Tetrabenazine prior to and at start of study | 28 |
| Table 10 TBZ 103,005: Baseline Demographic and Disease Characteristics..... | 28 |
| Table 11 TBZ 103, 005: Analysis of Change from Baseline in Max Chorea scores at Day 3..... | 29 |
| Table 12 TBZ103,005: Day 1 Mean Total Chorea Scores and Changes from Day 1 at Days 3 and 5 | 30 |
| Table 13 TBZ 103,004: Change in Chorea Score by Gender | 32 |
| Table 14 TBZ 103,004: Mean CGI by Gender | 32 |
| Table 15 TBZ 103,004: Change in Chorea scores by Age Group | 33 |
| Table 16 TBZ 103, 004 Mean CGI scores at Week 12 by Age Group | 33 |
| Table 17 Week 12 (or LOCF) Responses on Individual Items of UHDRS Part IV Functional Assessment Checklist | 37 |

LIST OF FIGURES

| | |
|----------------------------------------------------------------------------------------------------|----|
| Figure 1 TBZ 103,004: Change in UHDRS Max Chorea Score over Time (FDA Reviewer's Analysis)..... | 14 |
| Figure 2 TBZ 103,004: Distributions of CGI Global Improvement score at Week 12..... | 18 |
| Figure 3 TBZ 103,004: Treatment Group Difference in Change in Maximal Chorea Score by Center | 25 |

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

In one study, a 12 week study of tetrabenazine for the acute treatment of chorea in patients with Huntington's disease, the primary endpoint data support the proposed indication ($p < 0.0001$). The result on the clinical global improvement endpoint was also statistically significant in favor of tetrabenazine. However, results on two other secondary endpoints related to other aspects of Huntington's disease were nominally statistically significant in favor of placebo and this is the only acute study in the application. In this study there was also one suicide in the drug group but none in the placebo group. It should be noted that there is a high prevalence of suicide in Huntington's disease and twice as many patients were randomized to the drug. On the other hand, there were 8 (15%) depression adverse events in the drug group and 0 in the placebo group, which is a nominally significant difference. The other study was a very small 5 day randomized staggered withdrawal study. Although the group that had Tetrabenazine withdrawn first had a numerically higher mean chorea score than the other groups, suggesting a return of chorea upon withdrawal, the p-value was not significant ($p = 0.078$).

1.2 Brief Overview of Clinical Studies

Two studies were undertaken to support this application. The first was Study 103,004 - A 12 week randomized, double-blind, placebo controlled, multi-center study of tetrabenazine for the treatment of Huntington's chorea (Tetra HD). The second was Study 103,005 - A 5 day randomized, double-blind, placebo-controlled, staggered withdrawal study in patients with Huntington's Disease treated with Tetrabenazine. Both of these studies were conducted in the United States. Study characteristics are shown in Table 1.

Table 1 Study Characteristics

| STUDY | NUMBER RANDOMIZED | DURATION | DOSE | LOCATION/CENTERS | DEMOGRAPHICS | PRIMARY ENDPOINT |
|-----------------|----------------------|----------|---------------------------------------------------------------------------------------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| TBZ 103, 004 | 30 Placebo 54 TBZ | 12 weeks | 25-100 mg/day flexible dose titration to patient's "best" dose up to 100 mg | U.S. 16 centers | 62% Female 94% Caucasian mean age: 49 Mean Chorea scores at Baseline were 14.9 for TBZ 15.2 for placebo | Change from baseline in UHDRS item 12- Maximal Chorea |

| STUDY | NUMBER RANDOMIZED | DURATION | DOSE | LOCATION/CENTERS | DEMOGRAPHICS | PRIMARY ENDPOINT |
|-----------------|----------------------------------------------------------------------------------------------------------------------|----------|--------------------------------------------------------------------------------------------------|------------------|---------------------------------------------------|----------------------------------------------------------------------------------|
| TBZ 103, 005 | 12 Plac/Plac 12 TBZ/Plac 6 TBZ/TBZ (staggered withdrawal: day 1, day 3, day 5, respectively) | 5 days | patient specific stable dose prior to study entry range: 12.5-150 mg/day | U.S. 1 center | 60% Female, and 93% Caucasian. mean Age: 57 | Change from baseline in UHDRS item 12- Maximal Chorea at day 3 |

1.3 Statistical Issues and Findings

The 12 week acute treatment study, TBZ 103,004, was positive on the primary endpoint, the difference between the baseline and the average of the week 9 and week 12 scores on Unified Huntington's Disease Rating Scale (UHDRS) item 12 (maximal chorea), $p=0.0001$. Because the pre-specified primary analysis of the second study, the staggered withdrawal study, was not significant at the 0.05 level ($p=0.078$) it is important to check for internal replication in the positive, acute study. Within all individual sites except one group differences in the primary endpoint favored tetrabenazine. None were nominally significant but there was limited power since all sites had 9 patients or less.

The sponsor specified four secondary endpoints and proposed a prioritized order for testing each of them at 0.05 as long as all prior tests were significant at 0.05. The statistically significant treatment difference on the CGI-Improvement, the secondary endpoint that the sponsor considered the highest priority, provides some internal replication of the primary result. However, the sponsor's pre-specified analysis of the change from baseline to maintenance (average of week 9 and week 12 scores) in the UHDRS motor subscale did not reveal a statistically significant group difference. The difference was in the right direction but the p-value was greater than 0.05 ($p=0.075$). Because of this insignificant result and the sponsor's conditional sequential testing procedure any differences on the secondary endpoints of lower priority can only be considered exploratory. However, it should be noted that on the third secondary endpoint of priority, the functional assessment checklist (part IV of the UHDRS), a small but nominally statistically significant difference favoring placebo was seen. The fourth and final secondary endpoint in the prioritized list was the UHDRS gait score. No difference was observed in the UHDRS gait score.

Some of the other endpoints that were of lower priority than the four mentioned above had results that were somewhat unexpected. In particular, the placebo group was nominally significantly better than the tetrabenazine group on the change from baseline to week 12 in the sum of the cognitive items (UHDRS part II). Looking at the cognitive items individually, one finds that the group difference on the Stroop Interference Test – Words surpassed the nominal level of significance ($p=0.012 < 0.05$), the Stroop Interference Test-Interference nearly did ($p=0.053$), and placebo was numerically but not significantly better than tetrabenazine on the three other cognitive items. There was no significant group difference in mean change from

baseline through the maintenance period in the behavioral assessments (UHDRS part III), the independence scale (UHDRS part V), or the functional capacity scale (UHDRS part VI) but all three numerically favored the placebo group. Thus, the secondary endpoints provide limited internal replication and raise questions about the drug's effect on non-chorea aspects of Huntington's disease.

Patients with HAMD scores > 15, a benchmark for depression, were excluded from the study. The average baseline HAMD score was 5.1 for placebo and 4.5 for Tetrabenazine. Eight of 54 (15%) tetrabenazine patients reported depression as an adverse event as compared to 0 of 30 placebo patients (two-sided exact test $p=0.046$). Sadly, one of the eight tetrabenazine patients actually committed suicide. This reviewer could not locate the sponsor's HAMD analysis results, but the sponsor reported that there was no group difference between the baseline HAMD score and the average of the week 9 and week 12 HAMD scores. However, this reviewer estimated the group mean change by ANCOVA to be 1.6 (+/- 0.5 S.E.) points smaller for placebo than Tetrabenazine ($p=0.003$). A nonparametric test yielded the same conclusion. Despite the apparent group difference, the average week 12 score was still only about 2.5 for placebo and 3.9 for Tetrabenazine, so neither group was depressed on average. However, this may be because many patients (60%) were using antidepressants concomitantly. Thus, although the average week 12 HAMD scores did not suggest depression the nominally significant group difference this reviewer found in the change in HAMD scores corroborates the observed increase in depression related adverse events in the tetrabenazine group.

After the 103,004 study was underway a protocol was introduced for videotaping patients at the end of treatment (week 12) and one week after the cessation of treatment (week 13). An expert in Huntington's disease was to determine chorea scores for the videotapes without knowing the treatment group of the patient or to which visit the tape corresponded. This was done to support the primary analysis because it was felt that the investigators might be unblinded by the side effects of the drug. While the data from the videotapes seems to support the primary analysis result only 23 (27%) patients had videotapes made. Some patients who should have been videotaped were not, therefore, within the videotaped subgroup the treatment groups may not be balanced with respect to important baseline characteristics. For this reason it is not clear that the observed group difference within the subgroup with videos is due to the treatment alone. Therefore, the video rating results do not seem to have added much to the primary analysis result.

Although the group differences in chorea scores in the randomized staggered withdrawal study (TBZ 103,005) favored the combined group of those withdrawn on day 3 or day 5 over the group withdrawn at day 1 the primary analysis did not reach statistical significance ($p=0.078$). Fewer patients were enrolled than originally planned (30 vs. 45) reducing the power of the study after it was determined that a smaller sample size would be adequate because, apparently, patients were reluctant to agree to be taken off the drug. An ambiguity in the protocol resulted in patients that were supposed to be withdrawn on day 3 after the morning efficacy assessment, receiving placebo instead of tetrabenazine just prior to the day 3 morning assessment. The sponsor reasoned that since this made the 3 groups ordered at day 3 with respect to time of withdrawal a trend analysis would be more appropriate than the pre-specified comparison. A post hoc trend analysis yielded an unadjusted p -value of 0.048 but this would not be significant after adjusting

for the other tests that were conducted. In fact, a trend analysis was not specified in the protocol and would not have made sense for the day 3 data if the study had been conducted as planned because groups 2 and 3 would have been treated identically up to day 3. Thus, the trend analysis is an attempt to save the study from not only an insignificant primary result but also the error in study conduct and in this sense is a more of a stretch than a typical post hoc analysis. Note that four patients took protocol prohibited neuroleptics throughout the study and if these patients are excluded from the analyses neither the pre-specified primary comparison or the post-hoc trend analysis is nominally significant.

2 INTRODUCTION

2.1 Overview

Tetrabenazine, a selective centrally acting monoamine depletor, was initially developed by Hoffmann-La Roche in the mid -1950s as an antipsychotic drug. While the drug never gained wide usage as a tranquilizer, it was reported, in several small placebo controlled crossover studies, to be effective for the treatment of chorea, notably chorea associated with Huntington's disease (HD) with response rates reportedly ranging from 70 to 90%. Tetrabenazine was first approved in the UK for the treatment of chorea in 1971 and has been available in several European countries for over 30 years. In the US, patients have been receiving tetrabenazine for several years under physician INDs. Previous placebo controlled studies of tetrabenazine were primarily crossover studies conducted in small numbers of patients treated for only short periods of time (usually less than four weeks). The following two new studies were conducted to support this application.

Study 103,004 - A 12 week randomized, double-blind, placebo controlled, multi-center study of tetrabenazine for the treatment of Huntington's chorea (Tetra HD)

Study 103,005 - A 5 day randomized, double-blind, placebo-controlled, staggered withdrawal study in patients with Huntington's Disease treated with Tetrabenazine.

2.2 Data Sources

The data for study TBZ 103,004 can be found at the following location:

\\CDSESUB1\21894\N_000\2005-09-23\m5\datasets\tbz103.004\listings

The UH.xpt dataset contains the Unified Huntington's disease rating scale scores including items 12a-12g, the maximal chorea scores, the sum of which constitutes the primary endpoint.

The data for the 5 day randomized staggered withdrawal study, TBZ 103,005, are located in the following directory.

\\CDSESUB1\21894\N_000\2005-09-23\m5\datasets\tbz103.005\listings

The UH.xpt dataset contains the Unified Huntington's disease rating scale scores including items 12a-12g, the maximal chorea scores, the sum of which constitutes the primary endpoint.

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study TBZ 103,004

The date of the first patient's enrollment was July 9, 2003 and the date the last patient completed the study was March 15, 2004.

3.1.1.1 Study Design

Objectives

Primary:

The primary objective of this study was to establish the absolute reduction in chorea on optimized doses of tetrabenazine and placebo.

Secondary:

The secondary objectives were to determine the mean and standard deviation of the optimal dose and the percentage of participants responding at each dose level.

Study Design

This was a multi-center, randomized, double-blind, placebo-controlled, study of the efficacy, tolerability, and safety of tetrabenazine (titrated to best dose) in two parallel unbalanced (2:1) groups of participants suffering from manifest Huntington's disease. Huntington's disease was to be confirmed by the characteristic movement disorder (chorea); a positive family history of HD; and blinded CAG analysis during the study. Data from any participant who proved to not have HD by genetic testing was to be censored for the purposes of the primary analyses (note: it turned out that all randomized patients were positive for HD on genetic testing). Duration of double-blind treatment was to be 12 weeks, preceded by a screening period of no more than 2 weeks, and followed by a visit one week after the end of double-blind treatment. A total of at least 72 participants were to be enrolled in the study, and randomized in a 2:1 ratio to tetrabenazine or placebo. Doses were to be titrated up in increments of 12.5 mg per week to the dose that was best in terms of the desired effect and the tolerability of side effects (sedation and/or parkinsonism were expected to be the dose-limiting side effects). The minimum daily dose was to be 12.5 mg. The maximum daily dose was to be 100 mg. Study drug was to be administered q.d. and b.i.d. at the lower dosages of 12.5 and 25 mg per day, respectively, and t.i.d. for all other dosages. By the end of 7 weeks patients were to be at their best dose and this dose was to be maintained for the final 5 weeks of the double blind treatment phase. During the double blind treatment period, participants were to return to the clinic for efficacy, tolerability, and safety evaluations at the ends of Weeks 1 (± 2 days), 3 (± 3 days), 5 (± 3 days), 7 (± 3 days), 9 (± 3 days), and 12 (± 3 days). In addition, participants were to be contacted by phone during weeks 2 and 11. Finally, participants were to return to the clinic for a follow-up visit one week (Week 13 ± 2 days) after stopping treatment.

3.1.1.2 Efficacy Measures

Unified Huntington's Disease Rating Scale (UHDRS)

The Maximal Chorea score which consists of the sum of UHDRS items 12a-12g is the primary endpoint. Each of items 12a-12g is scored between 0 (chorea absent) and 4 (marked/prolonged). Thus, the Maximal Chorea Total ranges from 0 to 28.

The protocol was amended after the study was partially completed to require videotaping of the patients at the final visit on study drug (visit 6/week12) and off study drug at visit 7 (week 13). The videotapes were to be blinded according to visit and an independent expert in Huntington's disease was to score the chorea. These data were to be used in support of the primary endpoint. Because the videotaping was not originally planned only 23 (27%) patients had it done.

Secondary Endpoints

- Clinical Global Impression of Change
- parts I, II, III, IV, V and VII of the UHDRS.

3.1.1.3 Statistical Methods and Sample size

In accordance with the intent-to-treat principle, all participants randomized were to be kept in their originally assigned treatment group for analysis. All participants with at least one post-treatment evaluation were to be included in the efficacy analysis.

The primary efficacy analysis was to use change scores from baseline in Total Maximal Chorea Score. Secondary efficacy analyses were to use change scores from baseline in CGI, item 1; categorical analysis for CGI, items 2 and 3; and total scores on the UHDRS Parts I, II, III, IV, V, and VII. Change scores were to be measured from baseline to the end of Week 12, i.e., before any participants began to washout. These change scores were to be analyzed by ANCOVA, adjusting for sites and baseline scores. Two sided tests were to be used in the efficacy analyses. Additional analysis was to be performed to adjust for any significant imbalance of baseline characteristics.

The statistical analysis plan stipulated that the primary outcome measure was the change in Total Maximal Chorea score from baseline to the maintenance period. The maintenance score is defined as the average of the Week 9 and Week 12 scores. If either of these scores was missing the maintenance score was defined as the available score. For patients with neither a week 9 nor a week 12 measurement, the last available assessment was to be used in place of the value during the maintenance phase. Sites with fewer than three patients were to be pooled into one "site" for the analysis.

For the objective of finding the most effective and tolerable dose (subsequently called best dose), and dose-related efficacy, the frequency distribution of the best doses for the treatment group

was to be tabulated. The mean and standard deviation for these doses was to be obtained. The change score in total chorea from baseline to the end of week 12 at these doses was to be examined, and any relationship between the best doses and the change scores was to be recorded. The best doses were also to be correlated to other variables such as plasma concentration and baseline severity of chorea.

Sample Size Calculation

In this study, participants were required to have a Total Maximal Chorea Score ≥ 10 at baseline. Power calculations are based on the results obtained in a previous HSG study (Intro-HD) in the subgroup of patients who had a baseline Total Maximal Chorea Score ≥ 10 . Treatment duration in Intro-HD was 12 weeks, identical to treatment duration in this study. Based on this data a total of 24 placebo patients and 48 TBZ patients would provide 80% power to detect a group difference of 2.7 points in the Total Maximal Chorea change scores, assuming a standard deviation of 3.5 and allowing for a drop out rate of 15%.

A data and safety monitoring committee (DSMC) consisting of two physicians familiar with clinical research and a biostatistician, all of whom are otherwise not involved in the conduct of the study were to meet by conference call to review the protocol prior to the first randomization, after 20 participants had completed the study and finally after 40 participants had completed the study. The DSMC was to be unblinded and to have responsibility for reviewing AE occurrences and advising the PI, Steering Committee, and Sponsor in the event that the study should have been terminated for considerations of safety.

3.1.1.4 Disposition of Patients

A total of 84 patients were enrolled in this study; 54 patients were randomized to tetrabenazine and 30 to placebo. All enrolled participants were included in the primary efficacy analysis. Five (9%) tetrabenazine patients and one (3%) placebo patient did not complete the 12 week double blind treatment phase. The tetrabenazine withdrawals were due to the following adverse events: suicide; fall complicated by subarachnoid hemorrhage and confusion; suicidal ideation/psychosis/paranoia; pre-existing mass diagnosed as breast cancer; akathisia. The placebo patient withdrew consent.

3.1.1.5 Patient Demographics

Eighty-four HD subjects were randomized to tetrabenazine or placebo in two parallel, unbalanced (2:1) groups. The mean age was 49 years and ages ranged between 25 and 77 years. Sixty two percent of patients were female and 94% were white. Demographic characteristics and baseline disease characteristics were comparable between the groups.

Table 2 TBZ103, 004: Baseline Demographic Characteristics and Baseline Efficacy Measures

| VARIABLE | STATISTIC/ LEVEL | TBZ | PLACEBO | ALL | P-VALUE |
|---------------------------|---------------------|-------------|-------------|-------------|---------|
| Age | Mean (SD) | 49.4 (12.3) | 48.7 (10.5) | 49.2 (11.7) | 0.807 |
| Age Group | N(%) < 50 | 27. (50.0) | 15 (50.0) | 42 (50.0) | 1.000 |
| Age Group | N(%) ≥ 50 | 27. (50.0) | 15 (50.0) | 42 (50.0) | 1.000 |
| Race | N(%) Native Am | 2. (3.7) | 0 (0.0) | 2 (2.4) | 0.398 |
| Race | N(%) Black | 1. (1.9) | 0 (0.0) | 1 (1.2) | 0.398 |
| Race | N(%) White | 50. (92.6) | 29 (96.7) | 79 (94.0) | 0.398 |
| Race | N(%) Multiple | 1. (1.9) | 0 (0.0) | 1 (1.2) | 0.398 |
| Race | N(%) Unknown | 0. (0.0) | 1 (3.3) | 1 (1.2) | 0.398 |
| Gender | N(%) Female | 33. (61.1) | 19 (63.3) | 52 (61.9) | 0.841 |
| Gender | N(%) Male | 21. (38.9) | 11 (36.7) | 32 (38.1) | 0.841 |
| UHDRS 12a-g Max Chorea | Mean (SD) | 14.7 (3.8) | 15.2 (4.4) | 14.9 (4.0) | 0.578 |
| CGI-Sev | N(%) 3 | 12. (22.2) | 10 (33.3) | 22 (26.2) | 0.614 |
| CGI-Sev | N(%) 4 | 32. (59.3) | 16 (53.3) | 48 (57.1) | 0.614 |
| CGI-Sev | N(%) 5 | 9. (16.7) | 3 (10.0) | 12 (14.3) | 0.614 |
| CGI-Sev | N(%) 6 | 1. (1.8) | 1 (3.3) | 2 (2.4) | 0.614 |
| CGI-Sev | Mean (SD) | 4.0 (0.7) | 3.8 (0.7) | 3.9 (0.7) | 0.361 |
| UHDRS Functional | Mean (SD) | 18.8 (4.4) | 19.6 (3.8) | 19.1 (4.2) | 0.381 |
| UHDRS Gait | Mean (SD) | 1.2 (0.6) | 1.0 (0.5) | 1.1 (0.6) | 0.154 |
| UHDRS Motor | Mean (SD) | 47.0 (16.7) | 44.8 (15.4) | 46.2 (16.2) | 0.548 |
| CAG1 | Mean (SD) | 44.9 (3.4) | 44.3 (3.7) | 44.7 (3.5) | 0.490 |
| CAG2 | Mean (SD) | 17.6 (3.6) | 18.8 (2.8) | 18.0 (3.3) | 0.108 |
| DiseaseDuration | Mean (SD) | 8.6 (4.7) | 7.4 (4.5) | 8.2 (4.6) | 0.254 |
| Father HD | N(%) 0 No | 30. (55.6) | 12 (40.0) | 42 (50.0) | 0.432 |
| | N(%) 1 Yes | 21. (38.9) | 12 (40.0) | 33 (39.3) | 0.432 |
| | N(%) Unknown | 3. (5.6) | 6 (20.0) | 9 (10.7) | 0.432 |
| Mother HD | N(%) 0 No | 21. (38.9) | 16 (53.3) | 37 (44.1) | 0.078 |
| | N(%) 1 Yes | 30. (55.6) | 10 (33.3) | 40 (47.6) | 0.078 |
| | N(%) Unknown | 3. (5.6) | 4 (13.3) | 7 (8.3) | 0.078 |
| Prior- suicattempt | N(%) 0 No | 53. (98.1) | 29 (96.7) | 82 (97.6) | 0.670 |
| Prior- suicattempt | N(%) 1 Yes | 1. (1.9) | 1 (3.3) | 2 (2.4) | 0.670 |
| Prior-suic. ideation | N(%) 0 No | 45. (83.3) | 28 (93.3) | 73 (86.9) | 0.193 |
| Prior-suic. ideation | N(%) 1 Yes | 9. (16.7) | 2 (6.7) | 11 (13.1) | 0.193 |

The tetrabenazine and placebo groups were generally comparable for baseline HD characteristics. Disease duration was comparable in both groups. Disease severity, as judged by the CGI part 1 was also comparable in both groups. More in the tetrabenazine group reported that their mother was affected than in the placebo group (56% vs. 33%) but more in the placebo group did not specify whether their mother was affected (13% vs. 6%) and if the unspecified ones in the placebo group were mostly mothers it could resolve the difference.

The treatment groups were comparable for baseline chorea and HD severity as measured by total scores on the primary and secondary efficacy measures at baseline.

The protocol specified that patients with a total 17-item HAM-D score greater than 15 were not to be enrolled in the study. The mean HAM-D at baseline was 4.5 for the Tetrabenazine group and 5.1 for the Placebo group. Thirty (56%) tetrabenazine patients and 20 (67%) placebo patients took an antidepressant concomitantly with the study treatment.

3.1.1.6 Sponsor's Results

Prior to unblinding it was decided in the data analysis plan (dated April 2, 2004) that centers with 3 patients or less would be pooled. This resulted in the pooling of centers 104, 123, and 151. The primary efficacy analysis was an ANCOVA of the difference between the average of the week 9 and week 12 total chorea scores (sum of UHDRS items 12a-12g) and the baseline total chorea score. The model was adjusted for centers and treatment groups, and baseline total chorea score was included as the covariate. If a subject was missing either the week 9 or week 12 chorea score then the available score was used for the maintenance score (i.e., the average of week 9 and week 12). If both week 9 and week 12 were missing the last available post-baseline assessment was used. All but six participants completed the 12-week treatment period.

As seen in Table 3, chorea scores for participants in the tetrabenazine group declined from baseline to the maintenance period by a mean of 5.0 units, while those in the placebo group declined by 1.5 units. The treatment effect of 3.5 units is highly significant ($p < 0.0001$). A pre-specified sensitivity analysis imputed one plus the worst week 9 or week 12 chorea score (which happens to be $27+1=28$) for the three participants with missing data, 2 in the tetrabenazine group and 1 in the placebo group. After this imputation the corresponding p-value was 0.0015 which is still significant and suggests that the three missing scores would likely have little impact on the results.

Table 3 TBZ 103,004: Primary Efficacy Analysis Adjusted Mean Change (\pm S.E.M) in Total Chorea Score

| MEAN CHANGED TOTAL CHOREA SCORE (UHDRS ITEM 12) | | |
|-------------------------------------------------|------------------|-------------------------|
| Tetrabenazine (N=54) | Placebo (N=30) | P-value based on ANCOVA |
| -5.04 \pm 0.49 | -1.52 \pm 0.67 | <0.0001 |

* Based on an ANCOVA model with effects for baseline chorea score, sites, and treatment group

Figure 1 shows the group mean changes in UHDRS Max Chorea score over time. A nominally significant group difference in the mean change in chorea scores was seen as early as week 3 but no claim on this time of first difference is possible since no such determination was planned or accounted for in the primary decision rule (i.e., no alpha was allocated for testing at times before week 12). Note that the placebo group worsened by about a point from week 9 to week 12, while the tetrabenazine group improved by about a point.

Figure 1 TBZ 103,004: Change in UHDRS Max Chorea Score over Time (FDA Reviewer's Analysis)

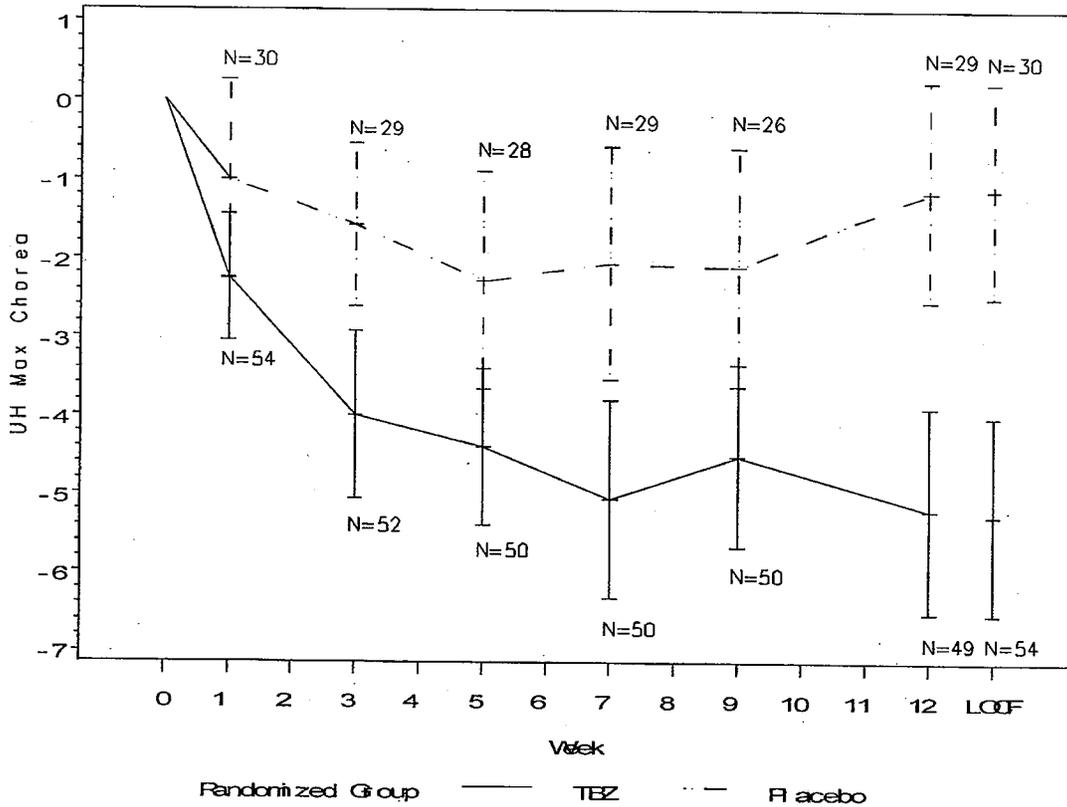


Table 4 shows the distribution of doses at week 7 (the end of titration) and the percent of patients at each dose that had a decrease in the chorea score of 3 points or more at the end of the study. Patients were not randomized to dose but rather titrated up to their "best dose". Doses of 50.0 and 100.0 were the most frequent in the tetrabenazine group. Ten of 11 (91%) at 50.0 mg had a 3 point decrease as compared to 13 of 22 (59%) at 100.0 mg. The highest dose might be expected to have a moderate proportion of non-responders because in the absence of AEs non-responders are titrated up in this study design.

In the placebo group almost all (93%) patients were taking the maximum number of placebo tablets (8). This suggests the possibility that if not all patients took the same number of tablets and one knew the number of tablets the patient was taking one could have guessed the treatment fairly accurately. Therefore, there could have been some unblinding of the investigators.

Table 4 TBZ 103,004: Dose at Week 7 and Percent with 3 point decrease from baseline in Total Chorea score

| | Treatment Name | | | | | | | | | | | |
|----------|----------------|------|------------------------------|------|-----|------|---------------|------|------------------------------|------|-------|------|
| | Placebo | | | | | | Tetrabenazine | | | | | |
| | N | Pct | 3pt Decrease in Chorea Score | | | | N | Pct | 3pt Decrease in Chorea Score | | | |
| | | | No | | Yes | | | | No | | Yes | |
| | | | N | Pct | N | Pct | | | N | Pct | N | Pct |
| wk7dose* | | | | | | | | | | | | |
| 0 | 0 | . | . | . | . | 1 | 1.9 | 1 | 100.0 | 0 | 0.0 | |
| 25 | 0 | . | . | . | . | 2 | 3.7 | 0 | 0.0 | 2 | 100.0 | |
| 37.5 | 0 | . | . | . | . | 5 | 9.3 | 3 | 60.0 | 2 | 40.0 | |
| 50 | 0 | . | . | . | . | 10 | 18.5 | 1 | 10.0 | 9 | 90.0 | |
| 62.5 | 0 | . | . | . | . | 3 | 5.6 | 1 | 33.3 | 2 | 66.7 | |
| 75 | 0 | . | . | . | . | 3 | 5.6 | 1 | 33.3 | 2 | 66.7 | |
| 87.5 | 2 | 6.7 | 1 | 50.0 | 1 | 50.0 | 8 | 14.8 | 2 | 25.0 | 6 | 75.0 |
| 100 | 28 | 93.3 | 22 | 78.6 | 6 | 21.4 | 22 | 40.7 | 8 | 36.4 | 14 | 63.6 |

* Actual dose is 0 for placebo but wk7dose/12.5 gives the number of placebo tablets taken

Withdrawal of Study Drug

At the end of week 12, study drug was discontinued per protocol and participants were followed up one week later. Participants on higher doses at the end of week 12, 5-8 tablets per day (62.5 to 100 mg), first had their dose reduced to 4 tablets at the start of week 13 and then discontinued entirely after 2 days on 4 tablets/day.

As seen in Table 5 below, week 13 mean Total Chorea Scores had returned to baseline levels in both treatment groups. The Tetrabenazine group average chorea score worsened by 5.67 points, a nominally significant increase, between week 12 and week 13. At week 13 the Tetrabenazine group average was 0.4 points worse than baseline while the placebo group average was 0.3 points better than baseline, but this difference is not statistically significant.

Table 5 TBZ 103,004: Mean Total Chorea Scores at Baseline, Week 12 and Week 13 (1 week after withdrawal)

| Treatment | BASELINE | | WEEK 12 (PRIMARY TIMEPOINT) | | WEEK 13(1 WEEK AFTER WITHDRAWAL) | |
|---------------|----------|--------------|-----------------------------|--------------|----------------------------------|--------------|
| | N | Mean ± SD | N | Mean ± SD | N | Mean ± SD |
| Tetrabenazine | 54 | 14.69 ± 3.84 | 54 | 9.41 ± 4.45 | 49 | 15.08 ± 4.21 |
| Placebo | 30 | 15.20 ± 4.41 | 30 | 14.07 ± 4.72 | 29 | 14.90 ± 4.47 |

3.1.1.7 Reviewer's Results

3.1.1.7.1 Primary Analysis

Centers 123 and 151 had less than 3 patients so they were pooled for the analysis as directed in the analysis plan. The maximal chorea score, UHDRS scale items 12a-g, ranges from 0 (best) to 28 (worst). The average baseline maximal chorea score was 15. As reported by the sponsor and verified by this reviewer, the primary endpoint, the difference between the average of the week 9 and week 12 scores and the baseline score, was estimated to be 3.5 (+/- 0.8 S.E.) points lower for the Tetrabenazine group. This is statistically significant, $p = 0.0001$. The change from baseline to week 12 (i.e., ignoring the week 9 score unless there was no week 12 score) was even larger (4.4 (+/- 0.9 S.E.) points, $p < 0.0001$) since the Tetrabenazine group mean decreased by a point between week 9 and 12, while the placebo group mean increased by a point during the same time.

3.1.1.7.2 Assessment of the Impact of Missing Data

In order to investigate the effects of missing data and adherence to the protocol specified visit times on the primary analysis result this reviewer defined several subgroups of the ITT population. This reviewer determined that 23 placebo and 40 Tetrabenazine patients had chorea scores for all scheduled visits and for which the last two visits fell within 1 ½ weeks of the protocol specified visit times. This subgroup is denoted OC₁. In this reviewer-defined observed cases subgroup the treatment group difference was slightly smaller than in the primary analysis but Tetrabenazine was still statistically significantly better than placebo ($p = 0.0013$). Another similar subgroup, denoted OC₂, dropped the requirement for having all visits prior to visit 6 but still required that visit 6 was within 1 ½ weeks of week 12. The result in this subgroup was also significant and close to the ITT-LOCF result. A pre-specified sensitivity analysis imputed one plus the worst week 9 or week 12 chorea score (which happens to be $27+1=28$) for the three

participants with no week 9 or 12 scores (2 in the tetrabenazine group and 1 in the placebo group). After this imputation the corresponding p-value was 0.0015. This reviewer found that if one was to impute the best possible score for the placebo patient and the worst possible score for the 2 TBZ patients with no week 9 or 12 scores then the results would still be significant although the p-value would increase by several orders of magnitude to 0.0155. This is still significant and suggests that the three missing scores would likely have little impact on the results. Since all of these sensitivity analyses still result in a significant result in favor of the tetrabenazine group the primary analysis result appears to be robust to dropouts and missing data as well as deviations from the protocol specified visit times.

Table 6 TBZ 103,004: Sensitivity Analyses of Primary Endpoint

| Population* | Placebo | | TBZ | | LSMean Difference | P-value |
|-------------------------------------------|---------|---------------------------|-----|---------------------------|-------------------|---------|
| | N | Change in UHDRS 12 LSMean | N | Change in UHDRS 12 LSMean | | |
| OC ₁ | 23 | -1.21 | 40 | -4.54 | 3.33 +/- 0.98 | 0.0013 |
| OC ₂ | 27 | -1.32 | 46 | -4.92 | 3.60 +/- 0.95 | 0.0003 |
| Sponsor's Worst Imputation | 30 | -0.83 | 54 | -4.25 | 3.42 +/- 1.04 | 0.0015 |
| Reviewer's Worst Case Scenario Imputation | 30 | -1.63 | 54 | -4.17 | 2.54 +/- 1.02 | 0.0155 |
| All-ITT-LOCF | 30 | -1.52 | 54 | -5.04 | 3.52 +/- 0.82 | 0.0001 |

* OC₁ had all scheduled visits and visits 5 and 6 were within 1.5 weeks of weeks 9 and 12 respectively

OC₂ had visit 6 within 1.5 weeks of week 12.

Sponsor's worst case imputation imputed 1 plus the worst observed week 9 or 12 score (27) for the 3 patients (2 TBZ and 1 placebo) that did not have a week 9 or 12 score.

Reviewer's worst case scenario imputation imputed 1 plus the worst observed week 9 or 12 score (27) for the 2 TBZ patients that did not have a week 9 or 12 score and the best possible score for the 1 placebo patient that did not have a week 9 or 12 score.

A few patients had visit 6, which was scheduled for week 12, considerably later. In particular, seven patients had visit 6 at week 14 or later including one at week 17. However, excluding the chorea scores from these late visits did not affect the significance of the primary analysis result.

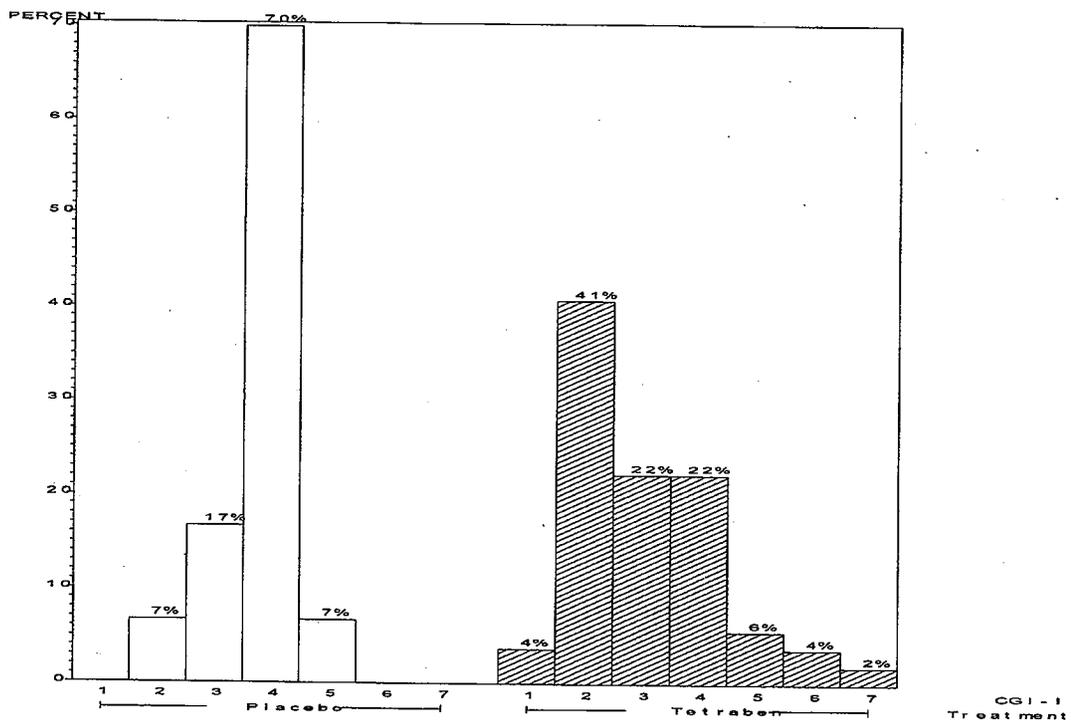
3.1.1.7.3 Secondary Analyses

The data analysis plan specified four secondary endpoints and proposed testing them conditionally in order (each at 0.05, given significance of previous endpoints at 0.05) to control the type I error rate. The specified order for testing was CGI Part 2 (Global Improvement),

UHDRS Total Motor Score, Functional checklist (sum of UHDRS questions 43-67), and, lastly, Gait score (UHDRS question 13).

The Tetrabenazine group was significantly better in terms of the clinical global improvement score as determined by the ANOVA analysis of the week 12 CGI score pre-specified by the sponsor ($p=0.005$). The estimated difference was 0.75 ± 0.26 (1 S.E.) points on the 7 point scale. The adjusted means were 3.75 for Placebo and 3.00 for Tetrabenazine. The CGI (part 2) score can only assume the integer values between 1 and 7 and, therefore, the appropriateness of using an analysis like ANOVA that depends on a normality assumption is questionable. Nevertheless, this reviewer found that the ANOVA result was corroborated by the nonparametric, center adjusted, Cochran-Mantel-Haenszel ANOVA test ($p=0.006$). Seven patients did not have the last visit on treatment, which was supposed to be at week 12, until week 14 or later. If we exclude CGI scores from these late visits and use the next to last post-baseline score instead for these patients, we obtain a slightly larger, but still significant, p -value of 0.019.

Figure 2 TBZ 103,004: Distributions of CGI Global Improvement score at Week 12



Part I of the UH scale is the motor assessment. The scores can range from 0 (best) to 124 (worst). Chorea items 12a-12g, the sum of which constitutes the primary endpoint, are part of the motor assessment. The average baseline score was 46. The mean difference between the average of the week 9 and 12 scores (the “maintenance score”) and the baseline score was estimated to be -3.51 for placebo and -6.84 for tetrabenazine. The estimated group difference is 3.3 (+/- 1.9 S.E.) points which is not nominally significant ($p=0.075$). Since this result is not significant at 0.05 the lower priority secondary endpoints, functional checklist and gait score, can not be tested without inflating the type I error. Note that the mean change from baseline to week 12 (not averaging over weeks 9 and 12) in the UH motor score (UH part I) was estimated to be -2.1 for placebo and -7.4 for Tetrabenazine, i.e., 5.3 (+/- 2.0 S.E.) points lower for the Tetrabenazine group. This is nominally significant, $p=0.012$. However, the analysis specified in the data analysis plan was the one described above that averaged over the week 9 and week 12 scores and did not produce a nominally significant result ($p=0.075$). Furthermore, although the sum of the non-chorea items of the motor assessment was not a secondary endpoint this reviewer investigated the possibility of group differences in the sum of the non-chorea motor assessment items. This sum can range from 0 to 96. The average baseline score was 30 for placebo and 32 for tetrabenazine. The group difference in the change from baseline to week 12 in the sum of the non-chorea items of the motor assessment was 1.5 +/- 1.5 (S.E.) points, numerically favoring tetrabenazine, but it did not reach the level of nominal significance ($p=0.32$). This suggests that the even if the observed difference on the change from baseline to week 12 in the motor assessment was significant it was primarily due to the difference on the chorea items.

Part IV of the UH scale is the functional assessment checklist. The scores can range from 0 (worst) to 25 (best). The average baseline score was 19. The average change from baseline to week 12 in the UH functional score (UH part IV) was 0.37 (slight improvement) for placebo and -0.81 for tetrabenazine (slight worsening). The difference was estimated to be 1.18 (+/- 0.49 S.E.) points lower (worse) for the Tetrabenazine group. This is nominally significant in favor of placebo, $p=0.018$. The difference between the average of the week 9 and 12 scores (the “maintenance score”) and the baseline score was almost identical for the functional assessment and was also significant in favor of placebo. A similar difference was also apparent at week 7 and week 9 but the clinical relevance of the group difference on the functional assessment is not clear since it is not large and analysis of individual items in the checklist did not reveal any significant differences. The exploratory p-value for item 52, which related to the ability to do laundry without help (TBZ: 70.4% able vs. Pla: 90.0% able, $p=0.051$), was the closest to reaching nominal significance. A table of the week 12 results for each individual item of the functional assessment checklist can be found in the appendix which starts on page 37.

It is important to note that at week 12 there was a difference in Item 68 of the UHDRS which identifies whether the patient or the patient and caregiver filled out the functional assessment checklist. More placebo patients filled out the checklist by themselves (47% vs 26% $p=0.04$). This may raise the question of whether the group difference may be attributable to the differences in who was filling out the checklist rather than the treatment. This is not a randomized subgroup so we can't be sure but the difference was still nominally significant in the larger subgroup of patients that filled out the checklist with their caregiver. The difference on item 68 was smaller and not significant at earlier weeks.

The gait score (UHDRS question 13) was the lowest priority of the four key secondary endpoints specified in the data analysis plan. Scores can range from 0 (normal) to 4 (cannot attempt). The baseline score was 1.0 for the placebo group and 1.2 for the tetrabenazine group. The average change from baseline to week 12 was 0.11 +/-0.06 (slight worsening) for placebo as compared to -0.03 +/- 0.06 (very slight improvement) for Tetrabenazine. This group difference was not significant ($p=0.241$) according to the analysis specified by the sponsor (ANCOVA). Since the gait score can only assume the integer values between 0 and 4 the appropriateness of using an analysis like ANOVA that depends on a normality assumption is quite questionable. A proportional odds logistic regression model (which simultaneously models for $j=0$ to 3 the odds of a response $\leq j$ as compared to a response $> j$) adjusting for site, baseline gait score, and treatment group yielded a p-value of 0.62. So, there was no apparent difference between the treatment groups in gait at week 12 (or early termination) as measured by the UHDRS item 13 score.

For a subgroup of 23 patients videos of the patients were made at weeks 12 and 13 and then rated by an independent specialist. The specialist was blinded to the patient's treatment and adverse events, as well as, the order in which the patient's videos were made. This procedure was only implemented after the study was already partially completed. The date of randomization of the first patient who had a video made was Oct 3, 2003. This reviewer found that only 44% of the patients randomized on or after that date had videos made at weeks 12 and 13. Two patients in site 55 and one patient in site 45 did not have videos made despite the fact that previous patients in their sites had had them made. This suggests that if there were different implementation times for the video protocol at different sites it still couldn't completely explain why some patients did not have videos. Since videos were not obtained from all patients after the protocol amendment requiring them, there may be imbalances between the treatment groups within the subgroup with videos. Thus, the apparent treatment group difference within the video subgroup could potentially have been influenced by imbalances in patient characteristics other than treatment. For example, the difference in average age between the groups is 11.5 +/- 6 (S.E.) years. In addition, the treatment group difference in the primary endpoint is estimated to be about 2.2 +/- 1.9 (S.E.) points larger in the subgroup with videos than in the subgroup without videos. The treatment group difference in unadjusted mean changes was 4.8 in the subgroup with videos as compared to 2.8 in the subgroup without videos. This suggests the possibility that the subgroup of patients with videos is not representative of the entire randomized population. For these reasons, although the video ratings appear to support the primary analysis the evidence is not without question.

Analyses of Other endpoints

The placebo group was nominally significantly better than the tetrabenazine group on the change from baseline to week 12 in the sum of the cognitive items (UHDRS part II). The placebo group improved by 5.1 +/- 4.5 points from an average baseline score of 172 +/- 55 whereas the tetrabenazine group worsened by 7.7 +/- 3.3 points from an average baseline score of 156 +/- 56. The estimated group difference was 12.8 +/- 5.6. Looking at the cognitive items individually, i.e., UHDRS items 19-23 in Table 8 below, one finds that the group difference on the Stroop Interference Test – Word Reading surpassed the nominal level of significance ($p=0.012 < 0.05$) and the Stroop Interference Test-Interference nearly did ($p=0.053$). Placebo was also numerically better than tetrabenazine on the three other cognitive items. There was no significant group difference in mean change from baseline through the maintenance period in the behavioral assessments (UHDRS part III), the independence scale (UHDRS part V), or the functional capacity scale (UHDRS part VI) but all three numerically favored the placebo group. Note that more positive scores are better on the cognitive items, the independence scale, and the functional capacity, whereas more negative scores are better on the behavioral items. These tests were exploratory and not adjusted for other comparisons but the fact that the placebo group was numerically better in so many cases and nominally significantly better in some is striking. Recall from above that the placebo group was nominally significantly better than the tetrabenazine group on the secondary endpoint functional assessment checklist (UHDRS Part IV) as well.

Part III of the UH scale is the behavioral assessment. Part III consists of 11 items each with two subitems a) frequency and b) severity. The subitems are scored from 0 (best) to 4 (worst). If we investigate the individual items that comprise part III we find that both anxiety items 27a (frequency $p=0.028$) and 27b (severity $p=0.040$) are nominally significant in favor of placebo (i.e, Tetrabenazine appears worse). At the end of week 12, 90% of placebo had no evidence of anxiety as compared to 70 percent of Tetrabenazine. At baseline there was no difference in these items ($p=0.56$ and $p=0.55$, respectively). Of course, these two comparisons were not adjusted for multiple comparisons, but these results might lead us to hypothesize that the drug is associated with the occurrence of anxiety. This hypothesis would need external validation. Note that anxiety was listed as an adverse event for 4/54 (7%) TBZ patients and 1/30 (3%) placebo patients. Anxiety aggravated was listed as an adverse event for 4/54 (7%) TBZ patients and 0/30 placebo patients. No other behavioral items had group differences that reached nominal significance.

Patients with HAMD scores > 15, a benchmark for depression in this study, were excluded from the study. The average baseline HAMD score was 5.1 for placebo and 4.5 for Tetrabenazine. Eight of 54 (15%) tetrabenazine patients reported depression as an adverse event as compared to 0 placebo patients (two-sided exact test $p=0.046$). Sadly, one of the eight tetrabenazine patients actually completed suicide. This reviewer could not locate the sponsor's HAMD analysis results, but the sponsor reported that there was no group difference between the baseline HAMD score and the average of the week 9 and week 12 HAMD scores. However, this reviewer found the group mean as estimated by ANCOVA to be 1.6 (+/- 0.5 S.E.) points smaller for placebo than Tetrabenazine. The group difference is nominally significant based on ANCOVA ($p=0.003$) or a Wilcoxon rank sum test (0.009) or a center adjusted Cochran Mantel Haenszel nonparametric ANOVA test ($p=0.029$). The last two tests are nonparametric tests which may be more reliable here since the scores are near the low end of the HAMD and thus the distribution may not be normal. Nevertheless, the various tests are in agreement. Despite the apparent group difference,

the average score was still only about 2.5 for placebo and 3.9 for Tetrabenazine so neither group was depressed on average. However, this may be because many patients (60%) were using antidepressants concomitantly. Thus, although the average week 12 HAMD scores did not suggest depression the nominally significant group difference this reviewer found in the change in HAMD scores corroborates the observed increase in depression related adverse events in the tetrabenazine group.

Note that no significant difference was observed on UHDRS item 25a - depressed mood frequency (a five point scale), UHDRS item 25b - depressed mood severity (a five point scale), or item 38-“does the examiner believe the participant is depressed?”. However, this doesn't seem to alleviate the increased incidence of depression adverse events in the tetrabenazine group or the significant group difference in the HAMD scores.

3.1.1.7.4 Summary of Secondary Endpoints Results

The sponsor designated four secondary endpoints as key secondaries in the data analysis plan. They planned to test them conditionally in the following order: CGI-I, UHDRS Motor, UHDRS Functional Assessment Checklist, and UHDRS Gait (item 13).

The only secondary endpoint that statistically significantly favored the tetrabenazine group was the first, the CGI-Improvement. The group difference on the 7 point scale was about 0.8 points ($p=0.005$). The group difference between the baseline score and the average of the week 9 and week 12 UHDRS motor total scores numerically favored tetrabenazine but was not significant ($p=0.075$), therefore, technically, testing should stop with this endpoint. It is important to note though that results on the third key secondary, the functional assessment checklist, were nominally significant in favor of placebo at week 12. A similar difference was also apparent at week 7 and week 9 but the clinical relevance of the group difference on the functional assessment is not clear since it is not large and analysis of individual items in the checklist did not reveal any significant differences. The exploratory p-value for item 52, which related to the ability to do laundry without help (TBZ: 70.4% able vs. Pla: 90.0% able, $p=0.051$), was the closest to reaching nominal significance. A table of the week 12 results for each individual item of the functional assessment checklist can be found in the appendix which starts on page 37. Differences on the other functional scales, the functional capacity (UHDRS part VI) and the functional impact scale, were not significant.

Part II of the UHDRS contains five items which measure cognitive abilities: Verbal fluency, Symbol digit modalities, Stroop Interference test -color naming, Stroop interference test - word reading, and Stroop interference test - interference. The result on the change in the sum of all five of the cognitive item responses was also nominally significant in favor of placebo as seen in Table 7 below. This may have been primarily due to the Stroop interference test items since there was a nominally significant difference in the change in the sum of the three Stroop items but not in the two non-Stroop items. However, all five items favored placebo numerically. In terms of individual items the word reading part of the Stroop interference test was nominally significant in favor of placebo (0.012) and the interference part of the Stroop interference test was nearly so ($p=0.053$). It should be mentioned that the p-values for the comparisons involving the cognitive

items were not adjusted for multiple comparisons. They were still felt to be important though since the placebo group was nominally significantly better than tetrabenazine in some cases.

Table 7 TBZ 103,004: Adjusted Mean Change from Baseline in UHDRS Part II - Cognitive Assessment Items

| SCALE OR ITEM | SCORING INFORMATION | PLACEBO (N=30) | | TBZ (N=54) | | DIFFERENCE LS MEAN (SE) | P-VALUE* |
|---------------------------------------------------|-------------------------------------------------------|--------------------|---------------------|--------------------|---------------------|-------------------------|--------------|
| | | BASELINE MEAN (SD) | CHANGE LS MEAN (SE) | BASELINE MEAN (SD) | CHANGE LS MEAN (SE) | | |
| Sum of Cognitive Items (Part II- items 19-23) | Higher scores are better Observed Range: 20-311 | 171.9 (55.2) | 5.1 (4.5) | 155.8 (56.2) | -7.7 (3.3) | 12.8 (5.6) | 0.025 (-) |
| Verbal Fluency (UHDRS 19) | Higher scores are better Observed Range: 2-51 | 19.0 (10.8) | -1.3 (1.1) | 18.9 (9.1) | -2.6 (0.8) | 1.3 (1.3) | 0.305 |
| Symbol Digit Modalities Test (UHDRS 20) | Higher scores are better Observed Range: 0-53 | 24.4 (11.5) | 3.0 (1.0) | 18.1 (11.5) | 2.1 (0.8) | 0.9 (1.3) | 0.509 |
| Stroop Interference Test (Color Naming; UHDRS 21) | Higher scores are better Observed Range: 8-82 | 47.2 (16.4) | 1.3 (1.8) | 42.4 (14.3) | -1.6 (1.3) | 2.9 (2.2) | 0.197 |
| Stroop Interference Test (wordreading ; UHDRS 22) | Higher scores are better; Observed Range: 0-110 | 56.5 (20.5) | 1.8 (2.1) | 53.8 (20.9) | -4.8 (1.5) | 6.6 (2.6) | 0.012 (-) |
| Stroop Interference Test (Interference; UHDRS 23) | Higher scores are better; Observed range: 0-56 | 24.7 (8.8) | 1.5 (1.2) | 22.6 (10.1) | -1.5 (0.9) | 3.0 (1.5) | 0.053 |

APPEARS THIS WAY
ON ORIGINAL

Table 8 shows the results for primary, key secondary and other endpoints. It includes information on the scoring of the endpoints and the baseline mean values and ranges in an effort to aid in the interpretation of the differences.

Table 8 TBZ 103,004: Adjusted Mean Change (Average Week 9 and Week 12) from Baseline in Endpoints

| SCALE OR ITEM | SCORING INFORMATION: | PLACEBO (N=30) | | TBZ (N=54) | | DIFFERENCE LS MEAN (SE) | P-VALUE* |
|----------------------------------------------------------------|-------------------------------------------------------------------|--------------------|---------------------|--------------------|---------------------|-------------------------|---------------|
| | | BASELINE MEAN (SD) | CHANGE LS MEAN (SE) | BASELINE MEAN (SD) | CHANGE LS MEAN (SE) | | |
| Max Chorea (Item 12a-g) -Primary Endpoint | Higher Scores are worse Possible Range: 0-28 | 15.2 (4.4) | -1.5 (0.7) | 14.7 (3.8) | -5.0 (0.5) | 3.5 (0.8) | <0.001 (+) |
| CGI-Improvement (Part 2) -Key Secondary #1 | Possible Range: 1 (Very Much Improved) to 7 (Very Much Worse) | N/A | 3.8 (0.2) | N/A | 3.0 (0.2) | 0.8 (0.3) | 0.005 (+) |
| Motor (Part I) -Key Secondary #2 | Higher scores are worse; Observed Range: 6-94 | 44.8 (15.4) | -3.5 (1.5) | 47.0 (16.7) | -6.8 (1.1) | 3.3 (1.8) | 0.075 |
| Functional Assessment Checklist (Part IV) -Key Secondary #3 | Higher scores are better; Possible Range: 0-25 | 19.6 (3.8) | 0.4 (0.4) | 18.8 (4.4) | -0.8 (0.3) | 1.2 (0.5) | 0.018 (-) |
| Gait (Item 13) -Key Secondary #4 | Scores can range from 0 (normal) to 4 (can't attempt) | 1.0 (0.5) | 0.1 (0.1) | 1.2 (0.6) | 0.0 (0.1) | 0.1 (0.1) | 0.241 |
| Sum of UHDRS Cognitive Items (items 19-23) | Higher scores are better Observed Range: 20-311 | 171.9 (55.2) | 5.1 (4.5) | 155.8 (56.2) | -7.7 (3.3) | 12.8 (5.6) | 0.025 (-) |
| Behavioral Assessments (Part III- items 25-35) | Higher scores are worse; Observed Range: 0-35 | 6.6 (6.2) | -2.2 (1.1) | 7.4 (7.3) | -1.0 (0.8) | -1.2 (1.4) | 0.363 |
| Independence (Part V Item 69) | Scores can range from 0 (max disabled) to 100 (not) | 80.2 (9.4) | 0.6 (1.3) | 76.9 (11.6) | -2.0 (1) | 2.5 (1.7) | 0.135 |
| Functional Capacity (Part VI Items 70-74) | Scores can range from 0 (max dysfunction) to 13 (none) | 8.6 (2.3) | -0.1 (0.3) | 8.3 (2.4) | -0.4 (0.2) | 0.4 (0.3) | 0.291 |
| Functional Impact Scale Total | Scores can range from 0 (independent) to 15 (complete assistance) | 0.41 (0.68) | 0.13 (0.23) | 1.30 (2.20) | 0.11 (0.17) | 0.01 (0.28) | 0.970 |

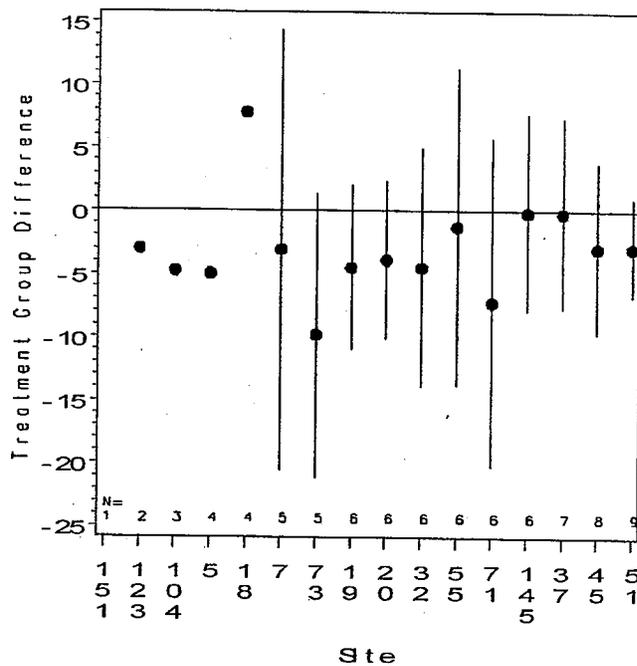
* based on ANCOVA model adjusting for baseline score, site, and treatment group

APPEARS THIS WAY
ON ORIGINAL

3.1.1.7.5 Group Differences on Change in Maximal Chorea within Individual Sites

Since there is only one acute study in this application we need to look for internal replication. One potential source of internal replication is significant treatment differences within individual centers. In this study there were 16 centers. As can be seen in Figure 3 below, almost all (14/15) of the treatment group differences on the change in chorea scores (difference between the baseline and the average of weeks 9 and 12) within individual centers favored Tetrabenazine, but none of the differences were nominally significant at the 0.05 level (based on an ANCOVA model including treatment and baseline as a covariate). In the figure the farther to the right a site is the more patients it had. The vertical lines passing through the black circles representing the means indicate the 95% confidence intervals. The vertical lines are not available for those sites on the extreme left because they had less than 2 patients in a treatment group. The failure of any intra-center treatment group differences to reach nominal statistical significance may be a case of a lack of power since the largest center had just 9 patients. Site 18, the only site with a treatment difference numerically favoring placebo, had just 1 placebo patient and 3 tetrabenazine patients.

Figure 3 TBZ 103,004: Treatment Group Difference in Change in Maximal Chorea Score by Center



3.1.2 Study TBZ 103,005

This study was initiated on November 11, 2003 and completed on December 10, 2004.

The primary objective of the study was to confirm the efficacy of tetrabenazine in Huntington's chorea by demonstrating in patients treated with tetrabenazine that when the drug is withdrawn chorea returns. A secondary objective was to evaluate whether chorea was more severe 5 days following treatment discontinuation than three days following treatment discontinuation.

3.1.2.1 Study Design

This was a single center, randomized, double-blind, placebo-controlled, staggered withdrawal study of tetrabenazine in three parallel unbalanced groups of participants suffering from manifest HD and being treated with tetrabenazine (administered at "best" dose). Duration of double-blind staggered withdrawal was to be no longer than 5 days. However, if during the double-blind portion of the study, participants experienced intolerable choreas (as judged by the Investigator), they could have been discontinued from the study. A total of at least 45 participants were to be enrolled in the study, 18(40%) were to initiate withdrawal on Study Day 1, 18 (40%) were to initiate withdrawal on Study Day 3, and 9 (20%) were to remain on tetrabenazine throughout the five-day study.

To be eligible, patients must have been receiving tetrabenazine for manifest HD as confirmed by clinical diagnosis and an expanded CAG repeat ($n \geq 37$). All patients were required to have been on a stable "best" dosage of tetrabenazine for two months prior to randomization and to have responded to this dose. Best dose is defined as the dose that provides moderate to marked improvement in the patient's condition while causing minimal side effects.

Participants were to be evaluated at screening, baseline/Randomization (Study Day 1), Study day 3, and Study Day 5. The duration of the study (five days) was justified on the basis of the plasma half-life of tetrabenazine (5.5 hours) and on published reports indicating that rapid return of chorea (within less than 24 hours), when tetrabenazine treatment is interrupted.

3.1.2.2 Efficacy Measures

The primary outcome measure is the change in Total Maximal Chorea Score (UHDRS questions 12a-g) from Baseline (Day 1) to Study Day 3.

Secondary efficacy parameters were to be the total score on the UHDRS Part I, II, III, IV, V, and VII. However, due to an administrative error (the protocol did not stipulate completion of these parts on day 3) these parts of the UHDRS were not collected at day 3, the timepoint for the primary analysis.

The functional capacity (part VI of the UHDRS) consists of three items scored between 0 (unable) and 3 (normal) and two items scored between 0 (unable) and 2 (normal). Thus the total score, sum of the 5 items, ranges from 0 to 13.

Reviewer's Comment: This assessment was listed under safety and tolerability assessments in the protocol rather than under efficacy assessments but in the study report the sponsor seems to regard it as a secondary efficacy endpoint.

3.1.2.3 Statistical Methods

Assuming a pooled standard deviation of the change score of the Total Maximum Chorea Score of 3.5, a sample size of 45 participants (18 in Group 1 starting withdrawal on Day 1; 18 in Group 2 starting withdrawal on Day 3; 9 in Group 3 with no withdrawal), would provide 80% power to detect a difference of 3.1 in change score between Group 1 and Group 2 + Group 3, using a two-sided level 0.05 test and allowing for a 10% drop-out rate.

The primary efficacy analysis was to compare Group 1 to Group 2 + Group 3 on change scores from baseline in Total Maximal Chorea Score on Study Day 3. Changed scores were to be analyzed by ANCOVA, adjusting for baseline scores.

Secondary efficacy analyses were to include:

ANCOVA analysis comparing group 1 to Group 2 + Group 3, on the change scores from baseline in total scores on the UHDRS Parts I, II, III, IV, V and VII on day 3.

Paired t-tests for Group 1 comparing changed scores in Total Maximal Chorea Score from Baseline to Day 3, and changed scores from Baseline to Day 5.

Exploratory analyses were to include ANCOVA analysis comparing the three treatment groups on the changed scores in all efficacy parameters from Baseline to Day 5.

3.1.2.4 Patient Disposition

There were major difficulties in enrolling patients into the study because patients already on tetrabenazine were reluctant to be withdrawn from tetrabenazine. A power calculation determined that the planned enrollment could be decreased from 45 to 30 participants without significantly compromising the ability of the study to detect treatment effects. Therefore, thirty patients were randomized (12 to Placebo/Placebo 12 to TBZ/Placebo, and 6 to TBZ/TBZ). All of them completed the study.

3.1.2.5 Patient Demographics

Since this is a withdrawal study patients were required to have been on stable doses of tetrabenazine for at least two months at baseline. The 30 randomized patients had been on tetrabenazine for an average of 2.5 years (the range was 0.21 to 7.07 years and the median time was 1.9 years). Summary statistics for the daily tetrabenazine dosage at study entry are displayed in Table 9.

Table 9 TBZ 103,005 Distribution of Stable Dose of Tetrabenazine prior to and at start of study

| GROUP | MEAN DAILY DOSAGE (MG/DAY) | MEDIAN DAILY DOSAGE (MG/DAY) | MINIMUM (MG/DAY) | MAXIMUM (MG/DAY) |
|----------------|----------------------------|------------------------------|------------------|------------------|
| Group 1 (N=12) | 59.38 ± 35.0 | 50.0 | 12.5 | 150 |
| Group 2 (N=12) | 45.83 ± 19.46 | 37.5 | 25.0 | 75.0 |
| Group 3 (N=6) | 54.17 ± 24.58 | 62.5 | 25.0 | 75.0 |
| All | 52.92 ± 27.4 | 50.0 | 12.5 | 150 |

Table 10 shows baseline demographic and disease characteristics of each group. Group 3, the group that stayed on tetrabenazine until day 5, was somewhat more affected at baseline than the other groups in terms of the CGI-Severity and the maximal Chorea score although the differences did not reach statistical significance. The randomized groups were reasonably comparable with respect to other characteristics.

Table 10 TBZ 103,005: Baseline Demographic and Disease Characteristics

| Variable | Levels | Pla/Pla | TBZ/Pla | TBZ/TBZ | All | Any Group Differences P-value |
|----------------------|-----------|------------|------------|-------------|-------------|-------------------------------|
| Age | Mean (SD) | 56.1 (9.7) | 55.9 (8.5) | 59.8 (14.2) | 56.8 (10.0) | 0.526 |
| Age Group | < 60 | 6 (50.0) | 8 (66.7) | 3 (50.0) | 17 (56.7) | 0.665 |
| Age Group | > 60 | 6 (50.0) | 4 (33.3) | 3 (50.0) | 13 (43.3) | 0.665 |
| Cgi-Sev | 3 | 0 (0.0) | 2 (16.7) | 0 (0.0) | 2 (6.7) | 0.170 |
| Cgi-Sev | 4 | 7 (58.3) | 7 (58.3) | 2 (33.3) | 16 (53.3) | 0.170 |
| Cgi-Sev | 5 | 4 (33.3) | 1 (8.3) | 1 (16.7) | 6 (20.0) | 0.170 |
| Cgi-Sev | 6 | 1 (8.3) | 2 (16.7) | 3 (50.0) | 6 (20.0) | 0.170 |
| Cgi-Sev | Mean (SD) | 4.5 (0.7) | 4.3 (1.0) | 5.2 (1.0) | 4.5 (0.9) | 0.262 |
| Disease Duration | Mean (SD) | 10.2 (4.5) | 9.2 (6.1) | 11.4 (4.8) | 10.0 (5.1) | 0.780 |
| Father HD | 0 No | 5 (41.7) | 5 (41.7) | 3 (50.0) | 13 (43.3) | 0.944 |
| Father HD | 1 Yes | 6 (50.0) | 7 (58.3) | 3 (50.0) | 16 (53.3) | 0.944 |
| Father HD | Unknown | 1 (8.3) | 0 (0.0) | 0 (0.0) | 1 (3.3) | 0.944 |
| Mother HD | 0 No | 6 (50.0) | 7 (58.3) | 3 (50.0) | 16 (53.3) | 0.944 |
| Mother HD | 1 Yes | 5 (41.7) | 5 (41.7) | 3 (50.0) | 13 (43.3) | 0.944 |
| Mother HD | Unknown | 1 (8.3) | 0 (0.0) | 0 (0.0) | 1 (3.3) | 0.944 |
| Prior suic attempt | 0 No | 12 (100.0) | 10 (83.3) | 6 (100.0) | 28 (93.3) | 0.200 |
| Prior suic attempt | 1 Yes | 0 (0.0) | 2 (16.7) | 0 (0.0) | 2 (6.7) | 0.200 |
| Prior suic ideation | 0 No | 11 (91.7) | 11 (91.7) | 6 (100.0) | 28 (93.3) | 0.765 |
| Prior suic ideation | 1 Yes | 1 (8.3) | 1 (8.3) | 0 (0.0) | 2 (6.7) | 0.765 |
| Race | Black | 0 (0.0) | 2 (16.7) | 0 (0.0) | 2 (6.7) | 0.200 |
| Race | White | 12 (100.0) | 10 (83.3) | 6 (100.0) | 28 (93.3) | 0.200 |
| Gender | Female | 7 (58.3) | 8 (66.7) | 3 (50.0) | 18 (60.0) | 0.784 |
| Gender | Male | 5 (41.7) | 4 (33.3) | 3 (50.0) | 12 (40.0) | 0.784 |
| Max Chorea (UHRS 12) | Mean (SD) | 9.4 (4.9) | 9.1 (6.2) | 11.2 (4.4) | 9.6 (5.3) | 0.594 |

3.1.2.6 Sponsor's Results

In Group 1 (Placebo/Placebo), mean Total Maximal Chorea Scores increased by 5.33 points between the Baseline visit and Day 3, and did not increase any further between Day 3 and Day 5, suggesting that wash-out was complete by Day 3. In Group 2(TBZ/Placebo), mean Total Chorea Scores increased by 3.6 points between the Baseline Visit and Day 3 and further increased by 1.9

points at Day 5. In Group 3 (TBZ/TBZ), mean Total Chorea Scores increased by 1.6 points between the Baseline Visit and Day 3 (this group remained on tetrabenazine between the Baseline visit and Day 3). In Group 3, mean Total Maximal Chorea Scores increased another 2.3 points between Day 3 and Day 5 following a 12-hour to 18-hour washout period.

As specified in the Data Analysis Plan, the primary outcome measure to be analyzed was the change in Total Maximal Chorea Score from the Baseline Visit to Day 3 where Group 1 (the group withdrawn from tetrabenazine at the Baseline Visit) was the experimental group and the combined Groups 2 and 3 (who should have received tetrabenazine prior to the Day 3 evaluations) was the control group. The mean change scores (\pm SD) from this analysis are summarized in Table 11. The Total Maximal Chorea scores for participants in Group 1 increased by a mean of 5.33 ± 3.47 units, while participants in the combined Groups 2 and 3 increased by a mean of 2.94 ± 3.52 units. The treatment effect was in the hypothesized direction with an estimated treatment effect of 2.39 units ($p=0.0779$).

The sponsor asserts that the withdrawal of tetrabenazine for group 2 on day 3 was to occur after the morning dose of the same study drug given on the previous day. However, because it was unclear in the protocol, the investigator made the switch before the morning dose on day 3 so that group 2 had been off tetrabenazine for 12-18 hours when the UHDRS maximal chorea ratings were made. Because this would tend to make group 2 more similar to group 1 the sponsor investigated several post-hoc analyses. The first investigated a difference between group 3 (TBZ/TBZ) and groups 1 (Pla/Pla) and 2 (TBZ/Pla) combined at day 3. However this difference was not nominally significant (Groups 1+ 2: 4.45 ± 3.20 vs. Group 3: 1.66 ± 4.71 , $p=0.1375$). A pairwise comparison between groups 1 and 3 also failed to reach nominal statistical significance (Group 1: 5.33 ± 3.47 vs. Group 3: 1.66 ± 4.71 , $p=0.1111$). The sponsor also performed a linear trend analysis, which they believe is reasonable because group 2 was withdrawn after group 1 but prior to group 3. This yielded a nominally significant result ($p=0.0486$) but this is also an exploratory post-hoc analysis. None of the planned primary or secondary efficacy analyses specified in the data analysis plan demonstrated nominally statistically significant treatment group differences.

Table 11 TBZ 103, 005: Analysis of Change from Baseline in Max Chorea scores at Day 3

| GROUP 1 (N=12) PLACEBO/PLACEBO MEAN \pm S.D. | GROUP 2+3 (N=18) TBZ/PLACEBO AND TBZ/TBZ MEAN \pm S.D. | ANCOVA P-VALUE FOR GROUP 1 VS. GROUP 2+3 |
|--------------------------------------------------------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------|
| 5.33 \pm 3.47 | 2.94 \pm 3.52 | 0.0779 |

3.1.2.7 Reviewer's Results

This study was conducted at a single site, Baylor College of Medicine. Because assessments for the primary analysis were taken shortly after withdrawal in this study there may have been a rebound effect, i.e., some patients may have had a transient dramatic worsening right after withdrawal that is not characteristic of the long term off-treatment efficacy score or the pre-treatment baseline score. However, since off-treatment baseline scores were not provided it is difficult to assess whether there was a rebound effect in this study. At day 3 a rebound effect could have affected the scores of group 1 (Placebo/Placebo), withdrawn on day 1, as well as group 2 (Placebo/TBZ), withdrawn on the morning of day 3.

This reviewer verified the sponsor's analyses and found that the p-value for the primary analysis was not significant ($p=0.078$). All patients completed the study so there are no missing data issues.

In addition to the primary comparison between the Placebo/Placebo group and the combined TBZ/Placebo and TBZ/TBZ groups at day 3, Table 12 shows p-values for other post-hoc comparisons. The post-hoc comparisons should be considered exploratory since they were not planned and are not adjusted for other tests.

Table 12 TBZ103,005: Day 1 Mean Total Chorea Scores and Changes from Day 1 at Days 3 and 5

| TREAT | Day 1 | | Day 3 | | | Day 5 | | |
|-----------------------|-------|-------------------------------|-------|--------------------------------|---------------------------------------------|-------|--------------------------------|---------------------------------------------|
| | N | Max Chorea Total MEAN (SD) | N | Change from Day 1 MEAN (SD) | P-value for comparison with Placebo/Placebo | N | Change from Day 1 MEAN (SD) | P-value for comparison with Placebo/Placebo |
| Placebo/Placebo | 12 | 9.4 (4.9) | 12 | 5.3 (3.5) | N/A | 12 | 5.3 (3.8) | N/A |
| TBZ/Placebo | 12 | 9.1 (6.2) | 12 | 3.6 (2.8) | 0.201 | 12 | 5.5 (3.4) | 0.918 |
| TBZ/TBZ | 6 | 11.2 (4.4) | 6 | 1.7 (4.7) | 0.062* | 6 | 4.0 (3.0) | 0.490 |
| TBZ/Placebo & TBZ/TBZ | 18 | 9.8 (5.6) | 18 | 2.9 (3.5) | 0.078 | | | N/A |

*based on an ANCOVA model including all 3 groups; ANCOVA model based on only Placebo/Placebo and TBZ/TBZ gives a p value of 0.111

If the study had been designed to be considered a win if either the protocol specified ANCOVA or the trend analysis was significant then the significance level would have to have been 0.025 (or less if more analyses were considered) to avoid inflating the type I error. The post-hoc p-value for the linear trend analysis is larger than 0.025 and thus would not be significant after the multiplicity adjustment. Furthermore, the trend analysis was not specified as even a secondary or exploratory analysis and if we adjusted for other secondary analyses the significance level for the trend analysis would have to be even smaller than 0.025. If group 2 had not been accidentally withdrawn from the drug before the day 3 morning assessments, then a linear trend analysis would not have been proposed. A trend analysis of the day 3 data among groups 1, 2, and 3 would not have made sense if the study had been conducted as the sponsor intended because group 2 and group 3 would have had identical treatment up to day 3, in which case $\mu_2 = \mu_3$.

Therefore, there would be no reason to expect the means to be ordered $\mu_1 > \mu_2 > \mu_3$, as required by a monotone trend, or as required for a linear trend (e.g., $\mu_1 > \mu_2 = \mu_1 - \beta > \mu_3 = \mu_1 - 2\beta$). Although there was limited power for detecting a difference between groups 2 and 3 the pairwise comparison between them is not nominally significant which would suggest that pooling these groups for the analysis as planned in the protocol is not necessarily inappropriate. A test for any differences (heterogeneity) among the 3 separate group mean changes at day 3 yielded a p-value of 0.15. None of the planned primary or secondary efficacy analyses specified in the data analysis plan demonstrated nominally statistically significant treatment effects. Due to an administrative error, patients were not rated on the non-chorea items of the UHDRS at day 3 so it is not possible to examine whether or not there were any treatment group differences on the following secondary variables: UHDRS parts I (Motor), II (Cognitive), III (Behavioral), IV (Functional), V (Independence), and VII (Clinical Summary). The Clinical Global Impression of Improvement was not administered after day 1 either. Thus, there were no secondary ratings to lend support to the insignificant primary analysis result.

Four patients (1 Pla/Pla, 2 TBZ/Plac, and 1 TBZ/TBZ) took prohibited neuroleptic medications (2 fluphenazine, 1 haloperidol, and 1 quetiapine) throughout the study. Excluding these four patients from the primary analysis yields a p-value of 0.118 and excluding them from the post-hoc analysis for trend also yields a p-value that exceeds 0.05 ($p=0.0814$).

Other Endpoints

The functional capacity assessment (UHDRS part VI) was listed under safety and tolerability assessments in the protocol rather than under efficacy assessments but in the study report the sponsor seems to regard it as a secondary efficacy endpoint. Notably, none of the Placebo/Placebo patients had a change in their score between day 1 and day 3 whereas the average change was -0.38 points (a slight worsening) in the combined other groups. Thus, the comparison between Group 1 (Placebo/Placebo) and Groups 2 (TBZ/Placebo) and 3 (TBZ/TBZ) combined at day 3 favored the Placebo/Placebo group numerically but did not reach the 0.05 level of significance (ANCOVA $p=0.35$).

3.2 Evaluation of Safety

Safety is not evaluated in this review. Please see the clinical review(s) for the evaluation of safety.

**APPEARS THIS WAY
ON ORIGINAL**

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

This section contains this reviewer's summary statistics for gender, race, and age subgroups. The studies were not adequately powered to estimate treatment effects precisely in subgroups or to detect differences between subgroups. Since these subgroups were not part of the decision rule and no adjustments were made for multiple testing the following p-values should be regarded as exploratory.

Gender

About 38% of patients in the acute treatment study TBZ 103,004 were male. While the treatment group difference in mean change in chorea scores was numerically larger in males than females treatment group differences were nominally significant for both males and females. In contrast to the results on the chorea scores the treatment group difference on the Clinical Global Impression of Change (CGI) was numerically larger for females than males. Thus, overall there was no compelling or consistent evidence that the treatment group difference varied significantly with gender in the 12 week acute study (04).

Table 13 TBZ 103,004: Change in Chorea Score by Gender

| TREAT | Female | | | Male | | | All | |
|---------|--------|---------------|---------|------|---------------|---------|-----|---------------|
| | N | MEAN (SD) | P-value | N | MEAN (SD) | P-value | N | MEAN (SD) |
| Placebo | 19 | -2.0 (4.2) | . | 11 | -0.9 (1.7) | . | 30 | -1.6 (3.7) |
| TBZ | 33 | -4.6 (4.0) | 0.005 | 21 | -5.4 (4.6) | 0.012 | 54 | -4.9 (4.4) |

Table 14 TBZ 103,004: Mean CGI by Gender

| TREAT | Female | | | Male | | | All | |
|---------|--------|--------------|----------|------|--------------|----------|-----|--------------|
| | N | MEAN (SD) | P-value* | N | MEAN (SD) | P-value* | N | MEAN (SD) |
| Placebo | 19 | 3.7 (0.8) | . | 10 | 3.8 (0.4) | . | 29 | 3.7 (0.7) |
| TBZ | 32 | 2.8 (1.1) | 0.003 | 20 | 3.5 (1.5) | 0.777 | 52 | 3.1 (1.3) |

*P values based on ANOVA

Forty percent of the 30 patients in the staggered withdrawal study (TBZ 103,005) were male. Since there were 12 or fewer patients per group in Study 05 it was too small to permit meaningful estimates of gender specific treatment differences regarding the effects of withdrawal or differences between genders. Although the gender specific means are not shown because of the small numbers of patients in each group none of the treatment group differences reached the nominal significance level of 0.05.

Age

In the acute study (TBZ 103,004) ages ranged between 25 and 77 and the mean age was 49. The following table shows that there was only a slight difference between the treatment effects on the change in maximal chorea score in the AGE < 50 and Age >= 50 subgroups. In fact, there was no compelling evidence that the treatment group difference varied significantly with age. Note that only 7 (9%) patients (5 TBZ and 2 placebo) were 65 or older, so a meaningful analysis of patients over the age of 65 is not possible.

Table 15 TBZ 103,004: Change in Chorea scores by Age Group

| TREAT | Age < 50 | | | Age ≥ 50 | | | All | |
|---------------|----------|---------------|---------|----------|---------------|---------|-----|---------------|
| | N | MEAN (SD) | P-value | N | MEAN (SD) | P-value | N | MEAN |
| Placebo | 15 | -0.7 (2.6) | . | 15 | -2.5 (4.2) | . | 30 | -1.6 (3.7) |
| Tetrabenazine | 27 | -4.3 (4.2) | 0.004 | 27 | -5.5 (4.2) | 0.009 | 54 | -4.9 (4.3) |

Table 16 shows the mean CGI improvement scores at Week 12 for each group. There was no significant difference in the treatment effects within the two age groups.

Table 16 TBZ 103, 004 Mean CGI scores at Week 12 by Age Group

| TREAT | Age < 50 | | | Age ≥ 50 | | | All | |
|---------------|----------|--------------|---------|----------|--------------|---------|-----|--------------|
| | N | MEAN (SD) | P-value | N | MEAN (SD) | P-value | N | MEAN |
| Placebo | 14 | 3.8 (0.4) | . | 15 | 3.7 (0.9) | . | 29 | 3.7 (0.7) |
| Tetrabenazine | 26 | 3.2 (1.4) | 0.16 | 26 | 2.9 (1.2) | 0.049 | 52 | 3.1 (1.3) |

The mean age in the staggered withdrawal study was 57. Since there were 12 or fewer patients per group in Study 05 it was too small to permit meaningful estimates of any age group specific treatment differences regarding the effects of withdrawal or differences between age groups. Although the age-group specific means are not shown because of the small numbers of patients in each group, none of the treatment group differences for the Age ≥ 60 and Age < 60 groups reached the nominal significance level of 0.05.

Race

Since only 5 (6%) patients in study 103,004 were not white no meaningful analysis of race subgroups is possible. Likewise, no meaningful analysis of race is possible for the withdrawal study 103,005, since only 2 patients were not white.

4.2 Other Special/Subgroup Populations

No other special populations or subgroups were investigated.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Because the pre-specified primary analysis of the second study, the staggered withdrawal study, was not significant at the 0.05 level ($p=0.078$) it is important to check for internal replication in the positive, acute study. Within all individual sites except one group differences in the primary endpoint favored tetrabenazine. None were nominally significant but there was limited power since all sites had 9 patients or less.

The sponsor specified four secondary endpoints for the acute study (TBZ 103,004) and proposed a prioritized order for testing each of them at 0.05 as long as all prior tests were significant at 0.05. The statistically significant treatment difference on the CGI part 2 ($p=0.005$), the secondary endpoint that the sponsor considered the highest priority, provides some internal replication of the primary analysis result. The sponsor's pre-specified analysis of the change from baseline to maintenance in the UHDRS motor subscale, which contains the chorea items of the primary endpoint, did not reveal a statistically significant group difference. The difference was in the right direction but the p-value was greater than 0.05 ($p=0.08$). Because of this insignificant result and the sponsor's conditional sequential testing procedure any differences on the secondary endpoints of lower priority can only be considered exploratory. However, on the UHDRS functional assessment scale a small but nominally statistically significant difference favoring placebo was seen. The fourth and final secondary endpoint in the prioritized list was the UHDRS gait score. No difference was observed in the UHDRS gait score. Thus, the secondary endpoints provide limited internal replication.

Some of the other endpoints that were of lower priority than the four mentioned above had results that were somewhat unexpected. In particular, the placebo group was nominally significantly better than the tetrabenazine group on the change from baseline to week 12 in the sum of the cognitive items (UHDRS part II). Looking at the cognitive items individually, one finds that the group difference on the Stroop Interference Test –Words surpassed the nominal level of significance ($p=0.012 < 0.05$), the Stroop Interference Test-Interference nearly did ($p=0.053$), and Placebo was numerically but not significantly better than tetrabenazine on the three other cognitive items. There was no significant group difference in mean change from baseline through the maintenance period in the behavioral assessments (UHDRS part III), the independence scale (UHDRS part V), or the functional capacity scale (UHDRS part VI) but all three numerically favored the placebo group. Thus, the secondary endpoints provide limited internal replication and raise questions about the drug's effect on non-chorea aspects of Huntington's disease.

Patients with HAMD scores > 15 , a benchmark for depression, were excluded from the study. The average baseline HAMD score was 5.1 for placebo and 4.5 for Tetrabenazine. Eight of 54

(15%) tetrabenazine patients reported depression as an adverse event as compared to 0 of 30 placebo patients (two-sided exact test $p=0.046$). Sadly, one of the eight tetrabenazine patients actually completed suicide. This reviewer could not locate the sponsor's HAMD analysis results, but the sponsor reported that there was no group difference between the baseline HAMD score and the average of the week 9 and week 12 HAMD scores. However, this reviewer estimated the group mean change by ANCOVA to be 1.6 (+/- 0.5 S.E.) points smaller for placebo than Tetrabenazine ($p=0.003$). A nonparametric test yielded the same conclusion. Despite the apparent group difference, the average week 12 score was still only about 2.5 for placebo and 3.9 for Tetrabenazine, so neither group was depressed on average. However, this may be because many patients (60%) were using antidepressants concomitantly. Thus, although the average week 12 HAMD scores did not suggest depression the nominally significant group difference this reviewer found in the change in HAMD scores corroborates the observed increase in depression related adverse events in the tetrabenazine group.

After the 103,004 study was underway a protocol was introduced for videotaping patients at the end of treatment (week 12) and one week after the cessation of treatment (week 13). An expert in Huntington's disease was to determine chorea scores for the videotapes without knowing the treatment group of the patient or to which visit the tape corresponded. While the data from the videotapes seems to support the primary analysis result only 23 (27%) patients had videotapes made. Some patients who should have been videotaped were not, therefore, within the videotaped subgroup the treatment groups may not be balanced with respect to important baseline characteristics. For this reason it is not clear that the observed group difference within the subgroup with videos is due to the treatment alone. Therefore, the video rating results do not seem to have added much to the primary analysis result.

Although the group differences in chorea scores in the randomized staggered withdrawal study (TBZ 103,005) favored the combined group of those withdrawn on day 3 or day 5 over the group withdrawn at day 1 the primary analysis did not reach statistical significance ($p=0.078$). Fewer patients were enrolled than originally planned (30 vs. 45) after it was determined that a smaller sample size would be adequate because, apparently, patients were reluctant to agree to be taken off the drug. An ambiguity in the protocol resulted in patients that were supposed to be withdrawn on day 3 after the morning efficacy assessment, receiving placebo instead of tetrabenazine just prior to the day 3 morning assessment. The sponsor reasoned that since this made the 3 groups ordered at day 3 with respect to time of withdrawal a trend analysis would be more appropriate than the pre-specified comparison. A post-hoc trend analysis yielded an unadjusted p-value of 0.048 but this would not be significant after adjusting for the other tests that were conducted. In fact, a trend analysis was not specified in the protocol and would not have made sense for the day 3 data if the study had been conducted as planned because groups 2 and 3 would have been treated identically up to day 3. Thus, the trend analysis is an attempt to save the study from not only an insignificant primary result but also the error in study conduct and in this sense is a more of a stretch than a typical post hoc analysis. Note that four patients took protocol prohibited neuroleptics throughout the study and if these patients are excluded from the analyses neither the pre-specified primary comparison or the post-hoc trend analysis is nominally significant.

5.2 Conclusions and Recommendations

The primary endpoint data from the 12 week study of tetrabenazine for the treatment of acute chorea in patients with Huntington's disease support the proposed indication ($p < 0.0001$). The result on the clinical global improvement endpoint was also statistically significant in favor of tetrabenazine. However, results on two other secondary endpoints related to other aspects of Huntington's disease were nominally statistically significant in favor of placebo and this is the only acute study in the application. In this study there was also one suicide in the drug group but none in the placebo group. It should be noted that there is a high prevalence of suicide in Huntington's disease and twice as many patients were randomized to the drug. On the other hand, there were 8 (15%) depression adverse events in the drug group and 0 in the placebo group, which is a nominally significant difference. The other study was a very small 5 day randomized staggered withdrawal study. Although the group that had Tetrabenazine withdrawn first had a numerically higher mean chorea score than the other groups the p-value was not significant ($p = 0.078$).

**APPEARS THIS WAY
ON ORIGINAL**

Appendix – Individual Items of the Functional Assessment Checklist

Since there was a significant difference favoring placebo in the change from baseline to week 12 in the sum of the functional assessment checklist item responses (UHDRS items 43-67) this reviewer investigated the results on the individual items that comprise the functional checklist. Each item is answered either yes or no. Table 17 shows the results. The items are presented in the table sorted by the size of the group difference in percentages that answered yes at week 12. The p-values should be considered exploratory since the tests were not pre-planned or adjusted for other analyses.

It is important to note that at week 12 there was a difference in Item 68 of the UHDRS which identifies whether the patient or the patient and caregiver filled out the functional assessment checklist. More placebo patients filled out the checklist by themselves (47% vs 26% p=0.04). This may raise the question of whether the group difference may be attributable to the differences in who was filling out the checklist rather than the treatment. This is not a randomized subgroup so we can't be sure but the difference on the change from baseline to week 12 in the sum of all items was still nominally significant in the larger subgroup of patients that filled out the checklist with their caregiver. The difference on item 68 was smaller and not significant at earlier weeks.

There were six items that had group differences greater than 15% in the percentage of patients that were able to do the item. Note that there were group imbalances at baseline on some of these items although none were significant at the nominal level. Most of the differences on individual items at week 12 were less significant after adjusting for the baseline responses. Item 52, related to doing laundry, has a p value of 0.051 after adjusting for the baseline responses. This was the smallest baseline adjusted p-value among the individual functional checklist items.

Table 17 Week 12 (or LOCF) Responses on Individual Items of UHDRS Part IV Functional Assessment Checklist

| UHDRS FUNCTIONAL CHECKLIST ITEM | LEVELS | BASELINE | | | WEEK 12 OR LAST OBSERVATION | | | | |
|-----------------------------------|-------------|--------------|----------------|---------------|-----------------------------|----------------|--------------------|--------------------------|---------------------------|
| | | TBZ (N=54) | PLACEBO (N=30) | CHISQ P-VALUE | TBZ (N=54) | PLACEBO (N=30) | PERCENT DIFFERENCE | UNADJUSTED CHISQ P-VALUE | BASELINE ADJUSTED P-VALUE |
| 68 Obtained from Participant Only | N(%) YES | 18 (33.3) | 10 (33.3) | 1.000 | 14 (25.9) | 14 (46.7) | -20.8 % | 0.053 | 0.043 |
| 47 Shop for Groceries | N(%) YES | 36 (66.7) | 24 (80.0) | 0.195 | 28 (51.9) | 22 (73.3) | -21.4 % | 0.055 | 0.159 |
| 49 Supervise children | N(%) YES | 28 (51.9) | 20 (66.7) | 0.189 | 25 (46.3) | 20 (66.7) | -20.4 % | 0.073 | 0.227 |
| 52 Do Laundry | N(%) YES | 44 (81.5) | 26 (86.7) | 0.541 | 38 (70.4) | 27 (90.0) | -19.6 % | 0.039 | 0.051 |
| 51 Do Housework | N(%) YES | 35 (64.8) | 22 (73.3) | 0.423 | 31 (57.4) | 23 (76.7) | -19.3 % | 0.078 | 0.089 |
| 59 Public transport | N(%) YES | 37 (68.5) | 23 (76.7) | 0.428 | 35 (64.8) | 25 (83.3) | -18.5 % | 0.072 | 0.096 |
| 55 Take meds w/o help | N(%) YES | 44 (81.5) | 28 (93.3) | 0.137 | 42 (77.8) | 28 (93.3) | -15.5 % | 0.067 | 0.265 |
| 46 Manage Finances | N(%) YES | 16 (29.6) | 11 (36.7) | 0.508 | 12 (22.2) | 11 (36.7) | -14.5 % | 0.155 | 0.170 |
| 50 Operate Auto | N(%) YES | 20 (37.0) | 14 (46.7) | 0.389 | 19 (35.2) | 14 (46.7) | -11.5 % | 0.302 | 0.570 |
| 60 Walk in | N(%) | 48 | 27 | 0.875 | 43 | 27 | -10.4 % | 0.222 | 0.205 |

| UHDRS FUNCTIONAL CHECKLIST ITEM | LEVELS | BASELINE | | | WEEK 12 OR LAST OBSERVATION | | | | |
|--------------------------------------------|-------------|---------------|----------------|---------------|-----------------------------|----------------|--------------------|--------------------------|---------------------------|
| | | TBZ (N=54) | PLACEBO (N=30) | CHISQ P-VALUE | TBZ (N=54) | PLACEBO (N=30) | PERCENT DIFFERENCE | UNADJUSTED CHISQ P-VALUE | BASELINE ADJUSTED P-VALUE |
| neighborhood | YES | (88.9) | (90.0) | | (79.6) | (90.0) | | | |
| 58 Bathe self | N(%) YES | 49 (90.7) | 29 (96.7) | 0.312 | 49 (90.7) | 30 (100.0) | -9.3 % | 0.086 | 0.941 |
| 44 Engage in any gainful employment | N(%) YES | 11 (20.4) | 7 (23.3) | 0.751 | 10 (18.5) | 8 (26.7) | -8.2 % | 0.383 | 0.326 |
| 57 Dress self | N(%) YES | 46 (85.2) | 29 (96.7) | 0.103 | 48 (88.9) | 29 (96.7) | -7.8 % | 0.217 | 0.935 |
| 45 Engage in volunteer or non gainful work | N(%) YES | 30 (55.6) | 17 (56.7) | 0.922 | 31 (57.4) | 19 (63.3) | -5.9 % | 0.596 | 0.502 |
| 48 Handle purchase | N(%) YES | 49 (90.7) | 26 (86.7) | 0.563 | 46 (85.2) | 27 (90.0) | -4.8 % | 0.531 | 0.273 |
| 54 Use telephone | N(%) YES | 52 (96.3) | 28 (93.3) | 0.541 | 48 (88.9) | 28 (93.3) | -4.4 % | 0.506 | 0.380 |
| 56 Feed self | N(%) YES | 51 (94.4) | 30 (100.0) | 0.189 | 50 (92.6) | 29 (96.7) | -4.1 % | 0.450 | 0.892 |
| 63 Comb hair w/o help | N(%) YES | 54 (100.0) | 30 (100.0) | .. | 52 (96.3) | 30 (100.0) | -3.7 % | 0.286 | 0.953 |
| 53 Prepare meals | N(%) YES | 39 (72.2) | 20 (66.7) | 0.594 | 38 (70.4) | 22 (73.3) | -2.9 % | 0.773 | 0.378 |
| 43 Engage in accustomed gainful employment | N(%) YES | 6 (11.1) | 4 (13.3) | 0.763 | 4 (7.4) | 3 (10.0) | -2.6 % | 0.680 | 0.782 |
| 61 Walk w/o falling | N(%) YES | 51 (94.4) | 27 (90.0) | 0.449 | 49 (90.7) | 28 (93.3) | -2.6 % | 0.680 | 0.433 |
| 67 Care provided at home | N(%) YES | 54 (100.0) | 30 (100.0) | . | 53 (98.1) | 30 (100.0) | -1.9 % | 0.453 | 0.950 |
| 64 Transfer between chairs | N(%) YES | 54 (100.0) | 30 (100.0) | . | 54 (100.0) | 30 (100.0) | 0 % | 1.000 | 0.953 |
| 65 Get in/out of bed | N(%) YES | 54 (100.0) | 29 (96.7) | 0.177 | 54 (100.0) | 30 (100.0) | 0 % | 1.000 | 0.953 |
| 66 Use toilet | N(%) YES | 54 (100.0) | 29 (96.7) | 0.177 | 54 (100.0) | 30 (100.0) | 0 % | 1.000 | 0.953 |
| 62 Walk w/o help | N(%) YES | 53 (98.1) | 29 (96.7) | 0.670 | 54 (100.0) | 29 (96.7) | 3.3 % | 0.177 | 0.809 |

APPEARS THIS WAY
ON ORIGINAL

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

/s/

Tristan Massie
3/23/2006 09:42:49 AM
BIOMETRICS

Kun Jin
3/23/2006 09:51:17 AM
BIOMETRICS

Kooros Mahjooob
3/23/2006 10:15:53 AM
BIOMETRICS