

**CATIE Schizophrenia Study
Neurocognitive Data
Post-baseline Data Analysis Plan**

Version: Final
Date: November 4, 2005

Outcome Measures

The primary outcome measure is the standardized neurocognitive composite score (SAS database variable NEURO_S). Note for all outcomes, responses are standardized to the baseline data.

Secondary outcome measures are the 5 standardized domain scores:

- Processing speed (SPEED_S)
- Reasoning (REASON_S)
- Working memory (MEMORY_S)
- Verbal memory (VERBAL_S)
- Vigilance (VIGIL_S).

Tertiary outcome measures are the 7 items that comprise 3 of the domain scores. Note that the Verbal Memory and Vigilance domains are each comprised of one item, while the other 3 domains are the average of the following items:

Processing Speed Items:

- Average of COWAT and Category Instances standardized responses (MNEUR34)
- Grooved Pegboard standardized response (MNEUR10)
- WAISR Digit symbol standardized response (NEUR9_S)

Reasoning Items:

- Wisconsin Card Sorting Test standardized response (MNEUR13)
- WISC Mazes standardized response (NEUR5_S)

Working Memory Items:

- Computerized test of visiospatial working memory std. responses (NEU12_SR)
- Letter Number sequencing test standardized response (NEUR6_S)

Collection Times

Neurocognitive data were collected at Baseline, Month 2, Month 6, Month 18, and the 2nd Month of treatment after switching to Phase 2 (Phase 2 Month 2). The neurocognitive battery was not administered at the time of phase switching.

Time points to be Analyzed

Phase 1/1A

Treatments will be compared at the following analysis time points during Phase 1/1A:

- Primary: Change from Baseline to Month 2
- Secondary: Change from Baseline to Month 6
Change from Baseline to Month 18
- Tertiary: Change from Month 2 to Month 6
Change from Month 2 to Month 18
Change from Month 6 to Month 18

Once a patient has discontinued from Phase 1/1A, their subsequent neurocognition data will be excluded from the analyses of Phase 1/1A.

There will be no adjustment of statistical p-values for multiple time point assessments. Significance of statistical results will be interpreted only for the primary time point; change from baseline to month 2. Statistical testing for other time points will be interpreted as a descriptive tool to help quantify the difference between treatment groups.

Note: It was determined that a mixed model including all 3 time points was not of interest because:

- Month 18 data is available for only 26% of patients. Using a mixed model to estimate missing data based on assessments taken a year prior did not seem reasonable.
- An analysis of the 3 individual time points separately is easy to interpret
- Most of the treatment effect is expected in the first 2 months of treatment. In addition, the sample size will be approximately 50% larger when comparing this treatment to the six month cognitive treatment response. Also, at 6 months there is a differential drop-out rate between olanzapine and the other treatments, as seen in figure 2 of the NEJM CATIE article.

Phase 2

Phase 2 analyses will be limited to the change from the last available measurement in Phase 1 or 1B to Month 2 in Phase 2. Analyses will be limited to patients who had at least 2 months of treatment on their previous phase prior to Phase 2.

- For patients who did not go into phase 1B, this is defined as change from the latter available assessment of month 2 or month 6 in Phase 1 to the Month 2 assessment in Phase 2.

- For patients who went into phase 1B, this is defined as change from month 6 in Phase 1B to the Month 2 assessment in Phase 2. Any Phase 1B patient with no month 6 assessment, or who was not in Phase 1B for 2 months prior to the month 6 assessment are excluded from this analysis.

Phase 1/1A Analyses

Multiple Comparisons and Determining Statistical Significance

The false-positive Type I error rate will be controlled for multiple comparisons due to: (1) multiple outcome measures, (2) multiple treatment groups within each outcome.

- No adjustment will be made for multiple time points: statistical significance will be determined only for change from baseline to Month 2. Statistical testing for all other time points is descriptive in nature to help quantify the difference between treatment groups.
- The primary outcome, the neurocognitive composite score, will be evaluated for overall statistical significance at month 2 with $p=0.05$. Adjustment for multiple treatment group comparisons is described in the below sections, Primary Analysis and ziprasidone analysis.
- The 5 secondary outcomes, the domain scores, will be evaluated for overall statistical significance at month 2 with a Hochberg adjustment for 5 multiple outcomes, in which the outcome with the largest p-value will be evaluated for overall statistical significance with $p=0.05$, and the outcome with the smallest p-value will be evaluated for overall significance relative to $p=0.05/5=0.01$. Within each outcome, adjustment for multiple treatment group comparisons follow the methods described in the below sections, Primary Analysis and Ziprasidone analysis, except that all adjustments will be made relative to the overall significance level assigned to each parameter (ranging from 0.01 to 0.05), rather than 0.05. In summary, the secondary outcomes have 2 levels of adjustment for multiple comparisons, to adjust both for the number of domains and the number of treatment group comparisons.
- Statistical testing for tertiary outcomes and all testing for all outcomes in Phase 2 month 2 is descriptive in nature.

Primary Analysis

Treatment groups will be compared for change from baseline using an analysis of covariance (ANCOVA), adjusting for baseline score, whether the patient had required crisis stabilization in the 3 months prior to study entry, and TD status where applicable (entry into Phase 1A, which excluded perphenazine). The method of treatment comparisons will

be identical to that specified in the original analysis plan and used in the primary outcome paper, as follows:

The four treatments available at the beginning of the trial will be compared overall with a 3 degrees of freedom (df) test, excluding patients with TD (in Phase 1A). If the overall test is significant at $p \leq 0.05$, then the three second generation antipsychotics (olanzapine, quetiapine, and risperidone) will be compared to each other for all patients (including TD patients in Phase 1A) using step-down testing. If the 2 degrees of freedom test is significant at $p \leq 0.05$, then each pair-wise treatment comparison will be evaluated relative to 0.05. In addition, if the overall 3 df test is significant, then perphenazine will be compared to the 3 second generation antipsychotics (excluding TD patients in Phase 1A) using a Hochberg adjustment for multiple comparisons in which the largest pair-wise p-value is compared to 0.05 and the smallest is compared to $0.05/3=0.0167$.

The Hochberg adjustment for multiple comparisons (Hochberg 1988) is a less conservative refinement of the Bonferroni correction, and maintains the overall Type I error rate at 0.05 as follows:

- all 3 comparisons are significant if the largest p-value is ≤ 0.050 .
- the 2 strongest comparisons are significant if the 2nd smallest p-value ≤ 0.025 ($0.05/2$)
- the strongest comparison is significant if the smallest p-value is ≤ 0.0167 ($0.05/3$)

Ziprasidone Analysis

Since ziprasidone was added to the study after 40% of patients had been enrolled, comparisons with ziprasidone are limited to the subset of patients randomized after it became available (the ziprasidone cohort). Pair-wise comparisons of ziprasidone with each of the other treatments are adjusted for multiple comparisons using a Hochberg adjustment in which the largest pair-wise p-value is compared to 0.05 and the smallest p-value is compared to $0.05/4=0.0125$. The comparison of ziprasidone vs. perphenazine excludes TD patients in phase 1A.

Evaluation of Potential Baseline Covariates

An evaluation of potential baseline covariates will be conducted for the primary Phase 1/1A assessment (change from baseline to month 2 for the neurocognitive composite score), for all intent-to-treat patients. In general, each variable will be evaluated separately for its ability to predict the neurocognitive change with an ANOVA. For continuous measures, a Pearson correlation will quantify the relationship. However, to address multiple comparison issues, some related covariates will be tested for association together in one regression model. These covariates are grouped together in the table below. For the models of groupings, the overall F-test of all added covariates will be evaluated. The three covariates to be included in the primary model will also be included in all models for evaluating continuous covariates.

Covariates to be examined include the following:

Grouping	Covariate	Notes
Included in Primary Model	Baseline Neurocognitive composite score	
	Exacerbation status	Required crisis stabilization in the 3 months prior to study entry (EXACER in DEMO Data set)
	TD status	entry into Phase 1A, which excluded perphenazine
Tier 1	Investigator site	Sites with 15 or fewer patients are pooled based on type of site (SITEGRP in the SITES data set)
	Gender	(GENDER in DEMO data set)
	Race	white vs. non-white (RACE_C in the DEMO data set)
	Hispanic ethnicity	(HISPANIC in the DEMO data set)
Tier II	Age	(AGE in the DEMO data set)
	Site type	mixed, private practice, private non-profit, VA, university, state mental health (SITETYPE in the SITES data set)
	Duration of Illness	years since first prescribed an antipsychotic (include age in the model as well, so we look at effects in addition to age) (YRS_PRES in the DEMO data set)
	Education Status	Years of education, WRAT reading score at BL English is first language (Yes/No) (EDUC_YRS. In the DEMO data set and WRAT3 and NEUR1 in the NEURO data set)
	Baseline antipsychotics	Olanzapine, Risperidone, Quetiapine, combinations including O, Q, and R, Any antipsychotic excluding O, Q, R (broken down by atypical vs. not), no medication (OLZIO, QUETIO, RISPIO, DUALMED0, ALLOTHIO and NONE0 in the DEMO data set)
	Baseline symptom scores	Total PANSS, positive, negative, and general psychopathology CGI-S at baseline (B1_PANSS, B1_PNEG, B1_PPOS, B1_CGIS in the PANSS data set)
	Baseline substance Abuse	Baseline Drug use/abuse, baseline alcohol abuse, use as defined by substance abuse working group (DRUG3 and ALC3 in the ABUSE data set)
	Experience with the neurocognitive battery	Indication of whether patients have previously been tested with at least 3 of the 9 tests (COWAT and Category Instances are combined together (variables BAT02 and BAT03), and the Facial-emotion test (BAT07) and WRAT-3 (BAT01) are excluded (NEURTEST data set)
	Baseline movement scores	AIMS, Barnes, Simpson-Angus scores at baseline (BL_INDEX, BI_GCA, BI_EPS in AIMS data set) Taking anticholinergics at baseline (ANTICHOL in MED data set)
Measures of Depression or Anxiety	Calgary depression score at baseline (B1_CALG) Baseline depression diagnosis Taking antidepressants at baseline Baseline other anxiety disorder diagnosis Taking benzodiazepines at baseline	

As a supportive analysis, the primary treatment comparisons described above will be repeated with adjustment for important covariates. All Tier I or II baseline covariates found to have a Type III $p < 0.01$ from an individual ANOVA (or grouped regression/ANCOVA for grouped items) will be included in the treatment ANCOVA models, and then any Tier I or II covariate with a Type III p -value > 0.05 after adjustment for treatment and the other covariates will be excluded from the final supportive model using backwards elimination.

The interaction between treatment group and covariates in the primary model as well as Tier 1 covariates will also be investigated. Each interaction will be evaluated individually in the primary ANCOVA model.

Evaluation of Potential Post-Baseline Moderators

The relationship between change in the neurocog composite score from baseline to Month 2 and change in other parameters will be evaluated by calculating the Pearson correlation between the two change scores, for all patients with a month 2 assessment in Phase 1/1A. For those that are found correlated with a p -value of $p < 0.01$, primary treatment comparisons for Phase 1/1A will be repeated with the inclusion of the change score for the other outcome measures as covariates. The purpose of this analysis is to see if there are treatment differences in neurocognitive changes above and beyond changes in related outcomes. If many parameters are found to have a correlation with change in composite score, then a backwards elimination strategy will be used to limit the adjusted analysis to moderators with a Type III p -value > 0.05 .

Potential post-baseline moderators include the following:

PANSS change from baseline	Note: PANSS was measured at months 1 and 3, but not 2, unless when month 2 is a patient's end of phase visit. Evaluate PANSS total, positive, negative, and gen psych subscores - include change from baseline to month 1, and change from baseline to month 3 as two separate measures. The month 3 analysis will be limited to people with month 3 data available.
Movement disorders change from baseline	Change in AIMS, Barnes, Simpson-Angus from baseline to month 1 and baseline to month 3. (these are measured at the same time as the PANSS) Indicator for new anticholinergic added since baseline If possible, also investigate interaction between medications and movement scores.
Compliance	Average compliance based on pill count from Month 1 and Month 2.

Post-baseline Substance abuse	Alcohol dependence/abuse, drug dependence/abuse at month 3 (based only on self-report and AUS/DUS). Post-baseline substance abuse is not yet defined by the substance abuse working group, so this is a preliminary evaluation
-------------------------------	--

In addition, change from Baseline to Month 2 will be compared overall and by treatment group for the subgroup of patients who discontinued Phase 1/1A anytime between months 2-6 vs. those who were still in Phase 1/1A at month 6. The purpose of this analysis is to see if neurocognition changes at month 2 were different for those who discontinued the phase after month 2 compared to those who stayed on longer.

Phase 2 Analyses

The two pathways of Phase 2 (efficacy pathway which includes clozapine, and tolerability pathway which includes ziprasidone) will be analyzed separately. Analyses will focus on the composite score; neurocognitive domains and items will be analyzed if sample sizes are determined to be large enough.

For the tolerability pathway, since ziprasidone was added to the study after a substantial number of patients had been enrolled, all analyses will be limited to the subset of intent-to-treat patients randomized after it became available.

Within the two separate pathways, treatment groups will be compared for change from baseline using an analysis of covariance (ANCOVA), adjusting for the applicable “baseline” assessment (month 2 or 6 from Phase 1/1A/1B), whether the phase 1/1A/1B assessment was from month 2 or 6, the actual duration from this score to the month 2 assessment in phase 2, whether the patient had required crisis stabilization in the 3 months prior to study entry, TD status (entry into Phase 1A, which excluded perphenazine), and Phase 1B status (for patients originally randomized to perphenazine). The method of treatment comparisons will be identical to that specified in the original analysis plan and used in the primary Phase 2 outcome papers, as follows:

The four treatments will be compared overall with a 3 degrees of freedom (df) test. If the overall test is significant at $p \leq 0.05$, then clozapine (for the efficacy pathway) or ziprasidone (for the tolerability pathway) will be compared to the 3 other antipsychotics using a Hochberg adjustment for multiple comparisons in which the largest pair-wise p-value is compared to 0.05 and the smallest is compared to $0.05/3=0.0167$. In addition, if the overall 3 df test is significant, the three second generation antipsychotics (olanzapine, quetiapine, and risperidone) will be compared to each other using step-down testing. If the 2 degrees of freedom test is significant at $p \leq 0.05$, then each pair-wise treatment comparison will be evaluated relative to 0.05.

Additional covariates to be included in a supportive version of the Phase 2 ANCOVA model include the reason for discontinuation from the previous phase, the treatment received in the previous phase, and interactions between these and the phase 2 treatment.

Evaluation of Potential Baseline Covariates in Phase 2

An evaluation of potential baseline covariates will be conducted for the primary Phase 2 assessment (change from baseline to month 2 for the neurocognitive composite score), for patients in the ziprasidone cohort of the ziprasidone pathway. (sample size in the clozapine pathway will be too small). The purpose of this analysis is to evaluate predictors from the previous phase (1/1A or 1B). Potential covariates will initially be evaluated with correlations, ANOVAs, or ANCOVAs. Covariates to be examined include the following:

- previous treatment group
- reason for discontinuation from previous phase
- duration of previous phase

As a supportive analysis, the Phase 2 treatment comparisons described above will be repeated with adjustment for covariates found to have a p-value < 0.05 after adjustment for treatment and other covariates. Important covariates from the phase 1 analysis may also be included. The interaction between treatment group and the above covariates will also be investigated.

Relationship of Neurocognition with Time to Discontinuation

An analysis of time to discontinuation will evaluate if neurocognitive functioning is related to the time until patients discontinued Phase 1/1A.

First, the association between baseline neurocognitive scores and time to discontinuation will be evaluated by adding the baseline composite score and each of the 5 domain scores individually to the Cox proportional hazards regression models used in the primary analysis for time until all cause discontinuation, as well as time until discontinuation due to lack of efficacy. If more than one domain score is found significant at $p < 0.01$, then an additional model will include all significant scores, and any with a p-value > 0.05 after adjustment for treatment and the other covariates will be excluded from the final supportive model using backwards elimination.

Next, the association between change in neurocognitive scores and time to discontinuation will be evaluated by adding the baseline, month 2, month 6, and month 18 composite scores as time dependent covariates in the Cox proportional hazards models used in the primary analysis for time until all cause discontinuation. This analysis will be repeated for each of the 5 domain scores, and also for the composite and each domain score for time until discontinuation due to lack of efficacy.