



## Design considerations for handling dropouts in anti-depressant drug trials

Jinglin Zhong<sup>a,\*</sup>, Ni A. Khin<sup>b</sup>, Peiling Yang<sup>a</sup>

<sup>a</sup> Division of Biometrics I, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Building 21, Silver Spring, MD 20993, USA

<sup>b</sup> Division of Clinical Compliance Evaluation, Office of Scientific Investigations, Office of Compliance, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Building 51, Silver Spring, MD 20993, USA



### ARTICLE INFO

#### Keywords:

Major depressive disorder  
Trial duration  
Time to discontinuation  
Efficacy endpoint

### ABSTRACT

**Background:** In clinical trials, statistical analysis often requires certain assumptions about missing data for a valid statistical inference. If the dropout rate is high, a wrong assumption about the missing data may compromise the validity of statistical inferences.

**Purpose:** To mitigate the high dropout rates commonly observed in psychiatry clinical trials, we consider two design approaches for short-term controlled trials submitted in support of marketing applications for drug products for the major depressive disorder (MDD) indication: (1) shortening the trial duration and (2) treating time to treatment discontinuation as an alternative primary efficacy endpoint.

**Methods:** Subject-level efficacy data from 45 trials for drugs approved for an MDD indication between 1997 and 2014 were collected. We analyzed change from baseline in Hamilton Depression Rating Scale (HAM-D-17) total score using the mixed model repeated measures approach. We compared the least squares means and the 95% confidence intervals of the treatment effect among three different trial durations, 4, 6, and 8 weeks. We considered two definitions of discontinuation: (i) all-cause discontinuation, (ii) discontinuation due to lack of efficacy. We compared the two-sided log-rank p-values with the p-values from the protocol-specified primary analysis.

**Conclusions:** Our findings suggest that MDD trials in the acute setting may be shortened to 6 weeks provided that the treatment difference between drug and placebo on HAM-D-17 total score reaches approximately 2 units at Week 6. However, our exploratory analyses of available data do not support the use of time to treatment discontinuation as an alternative primary efficacy endpoint.

### 1. Introduction

In randomized clinical trials comparing an investigational treatment with a control, missing measurements of the outcome of interest may compromise the validity of the statistical inference or introduce bias in estimating the treatment effect. The concern increases if there is a large amount of missing data, which is the issue in psychiatric clinical trials, where the average dropout rates are about 33% and 45% for major depressive disorder (MDD) and schizophrenia, respectively, in the acute-phase setting [1,2,3]. Although many statistical approaches have been proposed [4,5,6] to handle missing data, they all require relatively strong assumptions about the reasons for the missing data. Because the assumptions are generally unverifiable, it becomes a challenge to evaluate the impact of the high dropout rates. To mitigate the high dropout rate problem, from a design consideration, two possible approaches have been suggested: 1) shortening trial duration to decrease the dropout rate; 2) using time to treatment discontinuation as an

alternative primary endpoint. In this paper, we explore the feasibility of these two approaches by evaluating their potential impacts on the assessment of efficacy.

Reports of rising placebo response, declining treatment effect, and substantial failure rates in psychiatric trials are of great concern to the scientific community, the drug development industry, and the US Food and Drug Administration (FDA). We previously conducted exploratory analyses of trial-level data to investigate these trends in antidepressant and antipsychotic trials [1,2]; however, these approaches have significant limitations with regard to identifying possible contributing factors. Other researchers have compiled trial-level [16] and subject-level [17] data and used model based approach to identify contributing study design factors. However, contradicting findings on trial durations were found based on these trial-level data and the subject-level data analyses. To address these ongoing concerns, in 2012 we developed a pilot subject-level database comprised of approximately 7800 subjects with MDD who were enrolled in 24 short-term randomized controlled

\* Corresponding author.

E-mail address: [jinglin.zhong@fda.hhs.gov](mailto:jinglin.zhong@fda.hhs.gov) (J. Zhong).

monotherapy trials submitted as part of New Drug Applications (NDAs) over the previous decade. Exploratory analyses of this pilot dataset were conducted to evaluate the effect of trial duration and the use of different endpoints on trial efficacy. Preliminary findings were presented at the New Clinical Drug Evaluation Unit annual meeting in May 2013 [7] and the Drug Information Association annual meeting in June 2013 [8]. We then continued to build our MDD database by adding more trials and refining the data elements. We believe that what we have learned from our exploration of these data can provide direction on how acute MDD trials should be conducted to support drug approval. We hope this effort will provide the evidence needed to select more clinically meaningful endpoints for the assessment of effectiveness. This paper is a detailed summary of our presentation at the Annual Joint Statistical Meeting in August 2015, which included our updated results regarding the question of shortening trial duration and using time to treatment discontinuation as an alternative primary endpoint [9].

## 2. Methods

### 2.1. Data collection

We collected subject-level efficacy data from trials submitted to the FDA in support of NDAs for drugs approved for an MDD indication between 1997 and 2014. Subject-level datasets were not available in our electronic archives for NDAs submitted before 1997. We limited our search to randomized, double-blind, placebo-controlled, and short-term MDD trials. We further limited the search to trials with at least 40 subjects in at least one treatment arm and with at least one dose known to be effective. Subjects enrolled in these trials were adults diagnosed with MDD, according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). Trial data used in this analysis came from studies in which all participating subjects provided informed consent; however, personally identifiable information of these participants were not part of the data collection in the regulatory submissions.

The search resulted in 45 trials for eight approved antidepressants. Because of the confidential commercial nature of the information, individual drug names are not provided. In MDD trials, the most commonly used rating scales were the Hamilton Depression Rating Scale (HAM-D) [10] and the Montgomery-Asberg Depression Rating Scale (MADRS) [11]. Both the HAM-D and MADRS scales have been evaluated extensively [12–15]. For these trials, the pre-specified primary efficacy measure was change from baseline to endpoint in HAM-D or MADRS total score. Specifically, the HAM-D-17 (17 items), HAM-D-24 (24 items), and HAM-D-28 (28 items) were used as the primary efficacy measure in 27, 4, and 1 trial(s), respectively, while the MADRS was used as the primary efficacy measure in 13 trials. The HAM-D and MADRS scales are highly correlated and have similar sensitivity in detecting antidepressant efficacy [1,13–15]. The HAM-D-17 total score was collected in 41 trials regardless of whether or not it was pre-specified as the primary efficacy measure. Therefore, this paper focuses on these 41 trials, using mean change from baseline in HAM-D-17 total score. Thirty of the trials were conducted in North America (US and Canada). The other 11 trials were conducted in multiple geographical regions.

The 41 trials in which the HAM-D-17 total score was collected included a total of 16,073 evaluable subjects, defined as subjects who received at least one dose of treatment (drug or placebo) and had a baseline and at least one post baseline HAM-D-17 measurement. Of these subjects, 10,407 were assigned to 82 drug (including active comparator) arms and 5666 were assigned to 41 placebo arms. The mean baseline HAM-D-17 total score was 23. The mean age was 43 years. Approximately 63% of the subjects were female and approximately 83% were white. Trial durations were 6 to 10 weeks – the majority (35) of the trials were 8 weeks in length, and the remainder were 6 weeks (3 trials), 9 weeks (2 trials), and 10 weeks (1 trial) in length.

### 2.2. Data analyses

To explore the feasibility of shortening trial duration, we first obtained descriptive summaries based on observed values, such as the observed drug/placebo responses at each post-baseline visit for each subject in each arm within each trial; we did not impute values for missing data. For a subject, the response at a visit was defined as the change from baseline in HAM-D-17 total score at the visit. For each drug arm in each trial, we also estimated the mean treatment effect, defined as the mean drug response minus the mean placebo response in the same trial at each visit. When deriving the 95% confidence interval (CI) of the treatment effect, we did not adjust for multiplicity for trials where there were multiple drug arms. This decision was made because we wished to separate the impact of the design structure with multiple arms from the impact of the trial duration.

In addition to descriptive summaries outlined above, we also analyzed the data using the mixed model repeated measures (MMRM) analysis approach. Although analysis of covariance (ANCOVA) with the last-observation-carried-forward (LOCF) imputation method was the protocol-specified primary analysis in most of the trials, the MMRM analysis approach has been the most common primary analysis in psychiatry trials in recent years because of the concerns with the single-value imputation approaches, as noted by the Panel on Handling Missing Data in Clinical Trials, National Research Council [6]. The MMRM analysis approach requires the mechanism of missing data to be *Missing At Random* (MAR), which essentially assumes that the probability of missingness depends only on the observed data before subjects dropped out.

To compare treatment effects among three different trial durations (4 weeks, 6 weeks, and 8 weeks), for each duration of interest, we excluded observations beyond the targeted duration from the MMRM analysis. For example, to estimate the treatment effects over the 6-week duration, we included only observations up to the Week 6 visit even if the protocol-specified primary endpoint was change from baseline to Week 8. Then we performed a MMRM analysis for each trial, with the change from baseline in HAM-D-17 total score as the dependent variable, the baseline HAM-D-17 total score as a covariate, treatment, time (visit), and the interaction of treatment and time as factors, and subject as a random effect. For the covariance structure, we used *unstructured* unless there was a convergence problem, in which case we used structured covariance in the order of *toep*, *ar(1)*, and *CS* to reach convergence. We obtained the least-squares (LS) mean and the 95% CI of the treatment effect (i.e., drug response – placebo response) for each drug arm at the end of each duration.

To explore time to discontinuation as a possible endpoint, we considered two definitions of discontinuation: (i) all-cause discontinuation, (ii) discontinuation due to lack of efficacy (LOE). In each trial, we compared each drug arm with placebo using log-rank test and obtained nominal p-values (i.e., multiplicity adjustment was not considered in trials where there were multiple drug arms). Then we compared the 2-sided log-rank p-values with the 2-sided p-values from the protocol-specified primary analysis. As a way to estimate the overall treatment effect (drug relative to placebo), we obtained the hazard ratio from Cox regression model (with treatment arm as the only factor) to explore the relationship between the hazard ratio estimate and the p-values derived from the protocol-specified primary analysis.

## 3. Results

Fig. 1 shows the number of subjects at each week after randomization. In most of the trials, data were collected weekly for the first 2 weeks, then biweekly thereafter. In only a few trials, data at Weeks 3, 5, and 7 were collected. This resulted in much smaller numbers of subjects at Weeks 3, 5, and 7 than at Weeks 1, 2, 4, 6, and 8.

Download English Version:

<https://daneshyari.com/en/article/8757585>

Download Persian Version:

<https://daneshyari.com/article/8757585>

[Daneshyari.com](https://daneshyari.com)